Improving the Atmospheric Studies of Hot Jupiters from the Ground

Robert Wells
School of Physics & Astronomy
University of Southampton

May 1, 2015

Abstract

The goal of this project is to advance observational techniques of extrasolar planet atmospheres using large ground-based telescopes. These advancements will allow us to improve our knowledge of the chemical compositions of exoplanetary atmospheres, as well as their atmospheric circulation and thermal inversion layers. In this work I focus on the study of the hot-Jupiter WASP-18b via transmission spectroscopy. I analysed spectra of WASP-18b collected during three transit epochs with the multi-object IMACS spectrograph at the Las Campanas Observatory in Chile. My results show that it is possible to obtain transmission spectra of exoplanets with this instrument with a resolution of 6.7Å under 1 arcsec of seeing and an average precision of 3.06% per bin, per transit. This precision is not enough to detect the atmosphere of WASP-18b, because of this planet's very high mass and therefore small atmospheric scale height. However, with the precision I obtain it will be possible to measure atmospheric features in less massive, Jupiter-sized exoplanets. This work is part of ACCESS (the Arizona-CfA-Católica Exoplanet Spectroscopy Survey), which is a project to create the first comprehensive database of optical exoplanet spectra, using ground-based facilities.

Contents

1 Introduction

Over the past decade, the field of exoplanets has grown immensely. Almost 2000 exoplanets have been discovered to date [1], going from close-orbiting, Jupiter-sized planets (hot-Jupiters) to those similar to Earth in mass and size, e.g. Kepler-78b [2]. We have already started to probe the atmospheres of some hot-Jupiters, mini-Neptunes, and super-Earths around bright stars with some exciting and some disappointing results (i.e. many of the planets observed appear to have high thick opaque clouds that prevent us from extracting any information from the atmospheric layers below those clouds [3][4]). However, studying the atmospheres of these worlds is an important stepping stone towards future studies of potentially habitable planets, once those are discovered orbiting around nearby stars, since the observational strategies and technical challenges will be the same.

Most of the spectral observations of exoplanet atmospheres to date have been done from space, with the Hubble Space Telescope [5][6] and although some results have been obtained from the ground [7][8][9], those results are few and more marginal. However, the potential of ground-based observations is tremendous, since not only can we observe a larger number of planets, given the many telescope facilities, but we can also observe planets orbiting fainter stars given the larger mirror apertures of ground-based telescopes.

In this first section of the thesis I will briefly describe the open question in the study of exoplanet atmospheres that my thesis focuses on, i.e. what causes the presence of inversion layers in the atmospheres of giant planets? I will then discuss the technique I used, called transmission spectroscopy. In this section I also describe briefly the ACCESS collaboration, which this work is part of, and the target I have studied, the hot-Jupiter WASP-18b.

1.1 Thermal Inversions

The temperature of the atmosphere of Earth (or any planet) decreases gradually with height above the surface. However, there are areas which have sharp increases in temperature; these are called thermal inversions. One of these in the Earth's atmosphere is caused by the ozone layer in the stratosphere, as seen in figure 1. In this region oxygen molecules are split into atoms by UV photons from the Sun and then these atoms combine with oxygen molecules to form O_3 (ozone). This process, show in equation 1, heats up the surrounding area causing an inversion. The second inversion on Earth in the thermosphere seen in figure 1 is due to intense radiation from the Sun breaking up molecules such as O_2 and N_2 into ions.

$$\begin{aligned} O_2 + \gamma &\rightarrow 2O \\ O + O_2 &\rightarrow O_3 \end{aligned} \tag{1}$$

Thermal inversions have been found in hot-Jupiter exoplanets much like the ozone layer on Earth [10][11]. However, the temperatures of these planets', typically of the order of 2000K in the planet's day-side (i.e. the side irradiated by the star), are much too high for thermal inversions to be caused by oxygen. These thermal inversions have been attributed to titanium oxide (TiO) and vanadium oxide (VO) in gas form which are strong absorbers

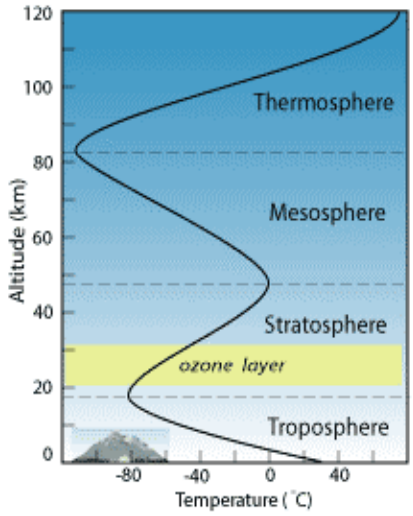


Figure 1: *Thermal inversions in Earth's atmosphere. The inversion in the stratosphere is caused by ozone and the inversion higher up is caused by other molecules. Image by Anne Egger (www.visionlearning.com/en/library/Earth-Science/6/The-Composition-of-Earths-Atmosphere/107).*

in the optical. They are theorised to be strong absorbers in the atmospheres of hot-Jupiters, by analogy to atmospheric processes in low-mass stars [12]. In this work I analyse transit observations of the hot Jupiter WASP-18b to investigate if we can achieve enough precision in the transmission spectra hot Jupiters to detect TiO.

In the following subsections I explain the technique used to measure the atmospheric spectrum of a transiting exoplanet.

1.2 Transmission Spectroscopy of Transiting Exoplanets

There are many methods for studying exoplanets including radial velocity, transits, timing variations, gravitational micro-lensing and direct imaging. Of these, only transits and direct imaging can be used to study atmospheric spectra. Direct imaging requires specific conditions to be viable (i.e. resolve the planet separate to the star), such as: the target system must be close to our Solar System to get high enough angular resolution to see the planet, planets need to be far from their host star and hot to be detected in IR which gives the best signal-to-noise. Although in the future direct imaging will become a very powerful technique for studying exoplanet atmospheres, to date only a few planets have been detected by this method, for example the HR 8799 system [13]. For transit events, the planet does not need to be resolved from it's parent star which makes this technique currently more useful for most exoplanet cases.

Transits although needing a chance orbital alignment for us to detect them, have been detected in large numbers and give the best way to gain information on a wide range of

exoplanets. Studying the atmosphere of an exoplanet at different wavelengths during transit is the best method to extract information about the atmosphere at different heights. According to the comprehensive catalogue exoplanet.eu [1], there are 1199 confirmed transiting exoplanets and a total of 1906 confirmed exoplanets from all techniques combined. Transits can even be studied from the ground with proper treatment of the systematics of the instrument and weather conditions.

Figure 2 shows a simple primary transit event with its corresponding light curve. A primary transit is the name given to the passage of the planet in front of its host star. When the planet starts to cross the star, the flux we receive begins to reduce; this is called the ingress and the reversed part at the end of the transit is called the egress. When the whole planet is in front of the star, there is a noticeable drop in flux over the course of the transit. An analogous event occurs when the planet crosses behind it's host star from our point of view. This is called a secondary eclipse with the drop in flux we observe corresponding to the loss in flux from the planet which is blocked by the star and is therefore less apparent than a primary transit. Secondary eclipses can be used for emission spectroscopy where the planet's emission spectrum is obtained. My work however, uses primary transits for transmission spectroscopy to obtain the planet's absorption spectrum. The depth of the (primary) transit (i.e. the difference in flux of inside and outside the transit) is related to the planet and stellar radius by equation 2. The arc of the light curve during the transit is due to limb darkening, so the centre of the star appears brighter than the limbs (edges) which means the transit depth is deeper towards the middle. This effect is larger at short wavelengths and gives a more rounded light curve. By fitting models to the shape of this light curve we can obtain various parameters of the planet, like the planet-to-star radius ratio, R_p/R_*, and atmospheric parameters, as aimed in this work.

The transit method gives a precise measurement of planetary radius, as seen by equation 2 where F_{out} and F_{in} are the fluxes in and out of transit, R_p is the planetary radius and R_* is the stellar radius which can be derived from models of stellar evolution or measured directly by asteroseismology [14]. By accurately measuring the flux from the system in and out of transit, one can obtain a precise measurement of the planets radius if the stellar radius is known well enough. Multi-object spectroscopy is an attractive way of studying transiting exoplanets by allowing the use of comparison stars to account for systematic effects of the data mainly due to the Earth's atmosphere, but also instrumental effects.

$$\Delta F = \frac{F_{out} - F_{in}}{F_{out}} = \frac{R_p^2}{R_*^2} \tag{2}$$

By using a method called 'transmission spectroscopy', we can observe variations in the planetary radius as a function of wavelength. In this method we do spectroscopy when the planet transits in front of its host star from our perspective and some of the light from the star passes through the planet's atmosphere. Depending on the chemical composition of the planet's atmosphere, photons of certain wavelengths are absorbed. At these wavelengths, the transits will appear deeper than at wavelengths with less or no absorption. These variations depend on the wavelength-dependent mean molecular absorption cross section, $\sigma_M(\lambda)$, and

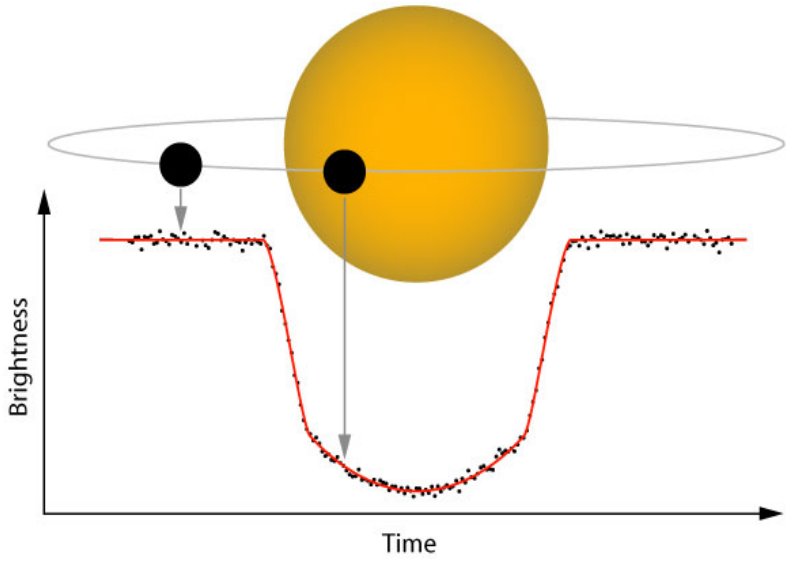


Figure 2: *The alignment of a transit and how a transit light curve looks at each moment in time. Image by the OSCAAR team (https://github.com/OSCAAR/OSCAAR/wiki/Introduction-to-Differential-Photometry).*

the atmospheric scale height,

$$h = \frac{k_B T}{m_H \mu_m g} \tag{3}$$

where k_B is the Boltzmann constant, T is the planet's mean atmospheric temperature, μ_m is the mean molecular mass, and g is the gravitational acceleration on the planetary optical surface. The scale height of an atmosphere is defined as the vertical distance above the surface over which the pressure or density drops by a factor of $1/e$, however in the case of gas giants (like WASP-18b), it is difficult to define a surface. A commonly adopted definition in gaseous planets is to define their 'surface' as the radius at which the atmospheric pressure is equal to 1 bar, which for Earth is about sea level. The mean molecular absorption is dependent on the planet's composition and is what is measured as variations in the transit depth at different wavelengths. Therefore by measuring changes with wavelength of the transit depth we can constrain both the physical and chemical structure of exoplanets; i.e. their temperature-pressure profile and chemical composition. This technique has had success already with detections of H_2O, CO and CH_4 in gas giants [15][16] and several planets with featureless spectra which point towards them having clouds and/or hazes [3][4].

All methods to obtain compositions of exoplanet atmospheres require theoretical models. Figure 3 shows the theoretical transmission spectra generated by ACCESS member Nikole Lewis for WASP-18b. There are 2 modelled spectra shown; a clear atmosphere (haze & cloud free) one which shows lines for sodium, potassium and water and one for which TiO is present. In the results section I will compare the values of WASP-18b I obtained to the theoretical spectral shapes to establish whether we can estimate its composition, or

otherwise, what the limitations are based on the observational error bars obtained. These observations are especially important from a theoretical standpoint because visible wavelength transmission spectra will be especially sensitive to the possible presence of TiO in exoplanetary atmospheres, which, as mentioned before, has been postulated as the source of thermal inversion in hot-Jupiter atmospheres [17]. Low mass stars show absorption lines of TiO in their spectra, which hints that a hot-Jupiter companion would have also contain some TiO. Such a detection would represent the first conclusive evidence of the presence of TiO in an exoplanet's atmosphere.

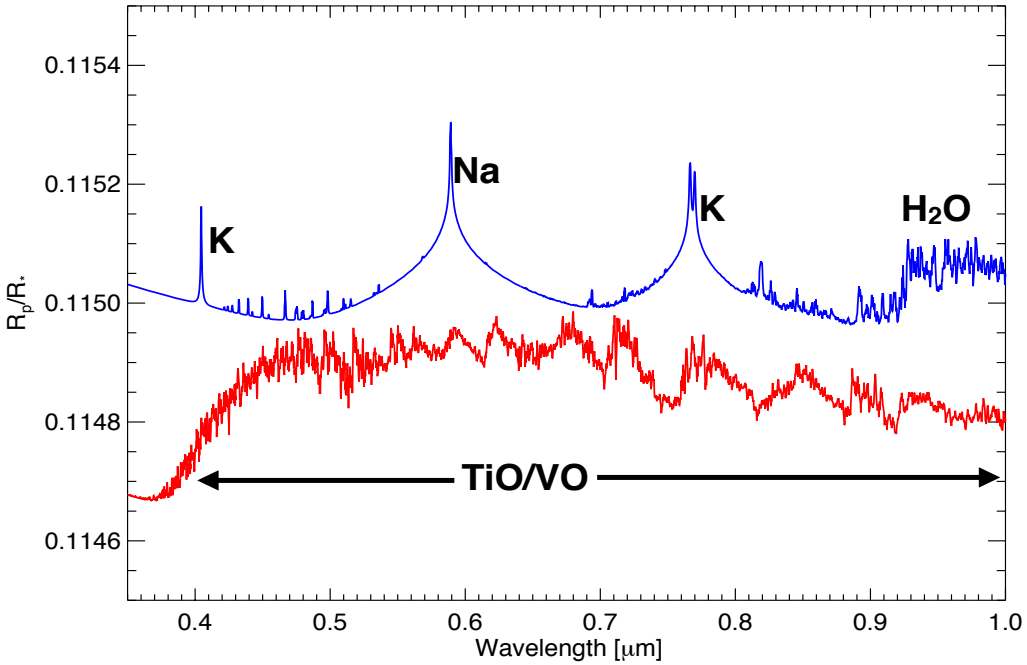


Figure 3: *Theoretical transmission spectra for WASP-18b. The blue plot is for a clear atmosphere with potassium, sodium and water included. The red line is the case if WASP-18b has TiO/VO present which could cause clouds but also be detectable in the transmission spectra.*

The interest in ground based transmission spectroscopy will soon increase greatly with the launches of upcoming space mission surveys like the Transiting Exoplanet Survey Satellite (TESS) [18], planned for 2017 by NASA. Over a 2 year period, TESS will survey the entire sky searching for transits on over 2 million nearby stars. It is expected to identify more than 1000 nearby transiting exoplanet candidates, many of which will be followed up from the ground and studied by transmission spectroscopy, making this technique hugely valuable in the near future.

1.3 The ACCESS Collaboration

This work is part of ACCESS (the Arizona-CfA-Católica Exoplanet Spectroscopy Survey), which is a project to create the first comprehensive database of optical exoplanet spectra, using ground based facilities. We use multi-object spectroscopy to collect simultaneous spectra from the target system and nearby comparison stars in the field of view. This allows us to account for systematic noise due to the Earth's atmosphere and the instrument. We repeat observations to validate our findings, commonly using 3 transits per target. The ACCESS sample, which can be seen in figure 4, covers a wide range of exoplanet parameters, with T_{eq} ∼500-2500K and R_p ∼2.5-24$R_\oplus$, which should probe a wide range of physical and chemical conditions. The target of this work, WASP-18b, has the largest equilibrium temperature in the sample and if detectable, its TiO content is expected to be large. The optical is the last uncharted region of exoplanetary transmission spectra, as the UV, near-IR and mid-IR have been more extensively explored from space. In this region, prominent lines of NaI and KI are expected, as seen in figure 3, as well as molecular bands and the visible (blue) continuum slope constrains the presence and size of haze particles and/or thermal inversion profiles. The collaborative effort of ACCESS aims to advance our knowledge about atmospheric processes, such as composition, cloud formation and thermal inversion layers.

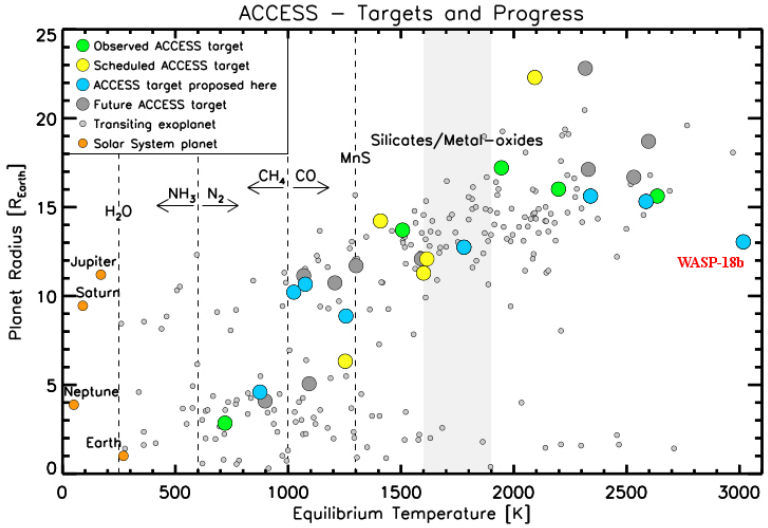


Figure 4: *Temperature-Radius distribution of known transiting exoplanets and Solar System planets. The large green, blue and yellow circles show, respectively, planets for which ACCESS already has at least one spectrum, planets for which we are currently requesting time, and the remaining planets in our sample. The vertical dashed lines and shaded area show the condensation temperatures for important atmospheric constituents. WASP-18b is the hottest planet in the ACCESS sample and is labelled in red.*

1.4 WASP-18b

WASP-18b was discovered in 2009 by Hellier et. al [19] and is still one of the most extreme exoplanets known due to it's high day-side temperature of about 2400K and close proximity to its host star, with a semi-major axis of 0.02AU. It is thought to have migrated in from where it formed further out and is now reaching the end of its spiral in towards it's host star [20]. The thermal profile of WASP-18b is theorised to have a thermal inversion which could not be confirmed or ruled out by Nymeyer et. al (2011) [21]. In their work they looked at secondary eclipses of WASP-18b using the Spitzer Space Telescope in 4 bands in the infrared between 3.6μm and 8.0μm. They found a blackbody model fits relatively well to the observed day-side brightness temperatures of the planet with a slightly better fit from an inversion model somewhat preferred over a non-inversion model. They also found WASP-18b to be much brighter than the predicted equilibrium temperature, requiring their models to use a near-zero Bond albedo and very low day-night energy redistribution ($\lesssim 0.1$). WASP-18b is one of the few exoplanets discovered which is hot enough that if there is a significant amount of TiO in its atmosphere then it will be in the gas phase and contribute to the transmission spectrum and therefore be detected. One potential problem is the small scale height of the planet, given its very large mass, which is not too far from the brown dwarf definition mass limit ($\sim$13M_J).

WASP-18 is a bright, near solar-mass star with a V band magnitude of about 9.3 [22]. This fact along with WASP-18b being a large planet with a radius of 1.106R_J means the transit depth is large and therefore gives good signal to noise. With it's short period of under a day it also transits often which makes it easier to obtain several datasets. The main properties of the system, including the planet and the host star are summarised in tables 1 and 2, with values taken from the discovery paper [19].

Property	Value	Uncertainty
Mass	10.30 M_J	$\pm$0.69 M_J
Period	0.94145299 days	$\pm$0.00000087 days
Radius	1.106 R_J	$\pm$0.072 R_J
Temperature	2384 K	$\pm$58 K
Semi-major axis	0.02026 AU	$\pm$0.00068 AU
Orbital inclination	86.0°	$\pm$2.5°
Eccentricity	0.0091	$\pm$0.0012
Transit length	2.144 hours	$\pm$0.016 hours

Table 1: *System properties of WASP-18b*

Property	Value	Uncertainty
Spectral type	F6	-
Mass	1.25 $M_\odot$	$\pm$0.13 $M_\odot$
Radius	1.216 $R_\odot$	$\pm$0.067 $R_\odot$
Effective temperature	6400 K	$\pm$100 K
Distance	100 pc	$\pm$10 pc
Age	1 Gyr	$\pm$0.5 Gyr
Metallicity, [Fe/H]	0.00	$\pm$0.09
Rotation, vsin(i)	11.0 kms^{-1}	$\pm$1.5 kms^{-1}

Table 2: *Stellar properties of WASP-18*

2 Observations

The data were collected on 3 separate nights between August 17 and November 7 2014, using the Inamori-Magellan Areal Camera & Spectrograph (IMACS) [23] in short camera mode (f/2) mounted on the 6.5m Magellan Baade telescope at Las Campanas Observatory, Chile. The 300 grism was used as the dispersing element, which gave a seeing-dependent spectral resolution, $\Delta\lambda$, of 6.7Å under 1.0 arcsec of seeing and a dispersion of 1.34Å per pixel. The nights were chosen because they were when transits of WASP-18b occurred. Details of the observations are shown in table 3. Each observation lasted for about 5 hours, taking between roughly 300 and 600 exposures and they were centred around times of the transits to get enough baseline before and after the transit to measure the depth well. More than one transit is needed to minimise potential systematics associated with a single dataset.

Dataset	**Date**	**Time (UT)**		**Exposures**		**Airmass**	**Conditions**
		Start	**End**	**Time (s)**	**Number**		
1	2014 Aug 17	03:45	08:32	5, 6	508	1.96→1.04	Good seeing
2	2014 Oct 20	04:26	08:48	1, 2	313	1.04→1.64	Moderate seeing
3	2014 Nov 7	00:52	06:17	2 - 6	607	1.20→1.04→1.28	Large flux variations at start due to clouds

Table 3: *Log of observations for WASP-18b*

The method of multi-object spectroscopy involves using a custom science mask on IMACS. This is a sheet of metal with holes drilled to only allow light through for certain stars in the field of view. IMACS in f/2 mode is very good for this sort of work due to it's large field of view of 27.20 arcmin x 27.20 arcmin. This allows for many stars in view to aid aligning the telescope, or to use as comparison stars which let us model out variations of the Earth's atmosphere. Such a wide field of view allows us to be more selective when choosing comparison stars as well; meaning we can pick stars similar to the target in both colour and magnitude. The mask contains 6 large (12 x 30 arcsec) slits for the science targets (WASP-18

and 5 comparisons) and 17 small slits (5 x 5 arcsec) for alignment. Such wide slits are used to eliminate slit losses and quantify the sky background. The mask can be seen in figure 5 along with how each spectrum looks on the 8 chips of the IMACS detector. The wide spectra correspond to WASP-18 and the comparison stars, and the narrow spectra correspond to the alignment stars which are not used for science.

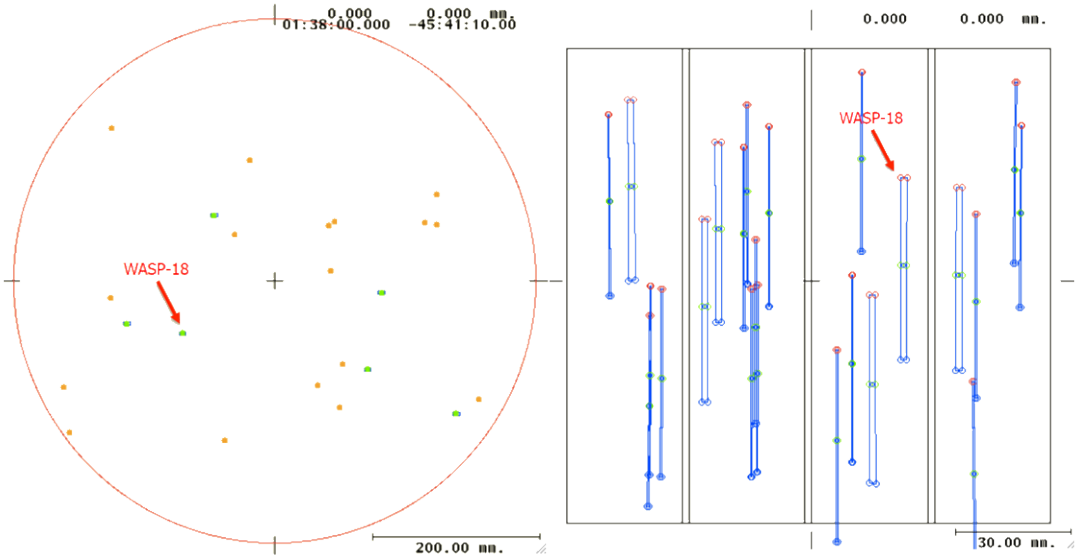


Figure 5: *Left: how the mask looks - green holes for WASP-18 & the comparison stars and orange holes for the alignment stars. Right: the layout of the spectra on the detector where the longer side is the spectral direction (wavelength space) and the other side is the spatial direction. There are 8 detector chips, in a 4x2 grid. The line splitting the 2 rows of 4 is not shown in the figure, but it causes each of the spectra to fall over 2 chips. This is mentioned in more detail later.*

Table 4 shows WASP-18 and each of the comparison stars in the field of view by their 2MASS [24] identifier, except for ob035 which could not be found in any catalogue. Also presented in the table is the name given to the star in the ACCESS pipeline used for data reduction and their apparent magnitude in V band. Of these, ob018 and ob035 were the only stars not to cover the entire wavelength range of interest (4000-10600Å), due to their spectra falling almost entirely on one chip and therefore there not being enough on the other chip to wavelength calibrate precisely because of less arc lamp lines.

In addition to the science mask, a second mask with the exact same number of slits, and of the same length but narrower (0.5 x 30 arcsec), was designed to observe wavelength calibration lamps. Arcs of a HeNeAr lamp taken using this mask and biases taken from the overscan region of the detectors, were later used to reduce the data and wavelength calibrate the spectra of the stars. The narrower slits in the calibration mask are necessary to produce distinct lines to wavelength calibrate the spectra which is explained in next section. Finally, the science mask (with wide slits) was used to collect spectroscopic flats using a Quartz lamp.

2MASS identifier	Pipeline name	V mag
2MASS J01372503-4540404	WASP18/ob001	9.342
2MASS J01383037-4537564	ob004	9.671
2MASS J01381558-4534146	ob009	10.779
2MASS J01383575-4529482	ob018	11.755
2MASS J01370966-4542309	ob029	12.305
not catalogued	ob035	12.701

Table 4: *The target and comparison stars used*

Quartz lamps lack any spectral lines, but produce instead a smooth and uniform illumination over the slits. The images of the uniformly illuminated slits on the detectors can be used to correct pixel sensitivity differences in the regions of the detectors where WASP-18 and the comparison stars fall. A total of 30 flat field images were collected before each of the three transit observations.

3 Data Reduction

Data reduction was done using a pipeline written in Python by the ACCESS collaboration. The main steps are described in this section.

3.1 Tracing and Sky & Background Subtraction

The pipeline took the raw .fits files from the observations, removed biases, created master flats by combining the individual flat observations of the night and applied flat field corrections to each of the science frames. The spectra in each image were initially traced by calculating the centroid of each row of pixels perpendicular to the dispersion (the spatial direction) and then fitting a 3rd order polynomial to them. Each row had 3 regions, first a central region which most of the light from the star. Either side of this was a middle region which contained sky continuum and an outer region which consisted of light outside of the slit, for example from scattered light. The outer region was used to find the sky background at each pixel along the dispersion direction. This was done by taking the median of 10 pixels both sides of the spectrum and fitting a 3rd order polynomial as a function of pixel. This smooth background varied slowly along the dispersion direction and was subtracted from the other 2 regions. Within the wide slits, the sky spectrum covered much more area on the detector than the stellar spectra, so the sky emission was identified row-by-row using the median of the spatial direction for each row. It is necessary to estimate the sky emission on a row-per-row basis because sky emission lines have a wide shape with sharp boundaries due to the fact they fully illuminate the slit. The sky emission was then removed from the central region, leaving only the stellar spectrum.

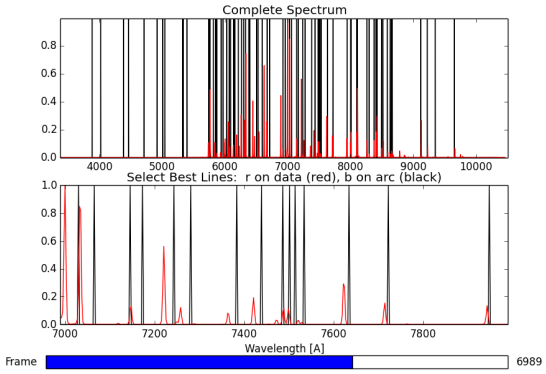
The sky and background subtracted spectra were then traced again more robustly. Each row of pixels (slice) was fitted with a Gaussian function to find the midpoint. The centres

of each slice were then traced using a 4th order polynomial. This also gave way to test the centres by comparing them to those obtained for the initial trace and check that both are consistent. The spectrum of each star was then simply extracted by summing 15 pixels either side of the trace for each row. This was wide enough to account for variations in seeing over the night, while comfortably staying inside the slit width of $\sim$50 pixels.

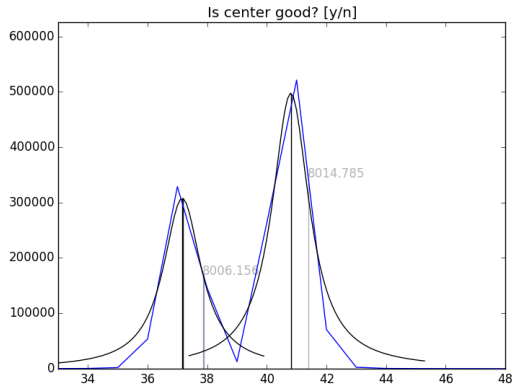
3.2 Wavelength Calibration

The arc lamp used contained helium, neon and argon which when excited emit light at precisely known wavelengths. Using 2 python scripts written by the collaboration, I was able to match the lines of the arc lamp spectra to the pixels on the chip for each star, based on the location of the star and arcs spectra on the detector.

I needed to switch from spectra against pixel to spectra against wavelength. To do this I compared the arcs to a data file with the known emission lines and fitted Lorentzian profiles to the lines of the arcs to get the line centres. The 'wavelength solution' of pixel to wavelength was then obtained by fitting a 6th order polynomial to the line centres. This process was iterated, removing the furthest outlying point each time until the root mean square error was less than 0.05Å. Between 60 and 68 of the 70 lines I had on file were used for each stars spectrum. The computed wavelength solutions were only used for the first science image of the night, the other images were cross-correlated with the first solution by fitting a 3rd order polynomial with time to account for smooth wavelength shifts and put on a common wavelength grid. Lastly, all the stars were calibrated to the same physical reference frame by finding the shifts between the H_α absorption line (6562.8Å) of the median spectra and the value of the line in a vacuum. In figure 6 I illustrate the interactive windows of the of the python wavelength calibration scripts used, first to guess the location of the lines and then to find the line centres.



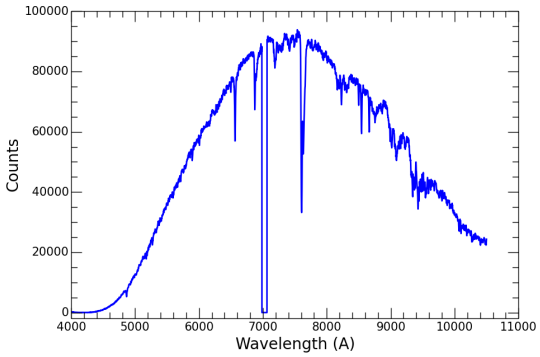
(a) The script to match the lines



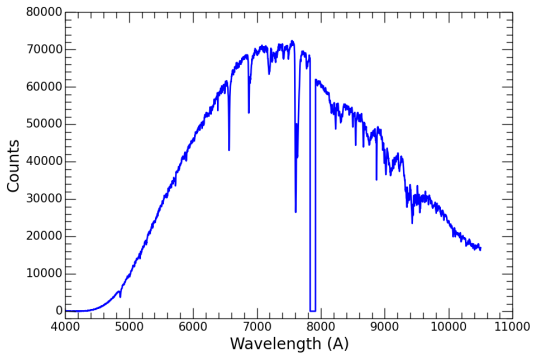
(b) The script to obtain line centres

Figure 6: *Examples from the scripts used to wavelength calibrate the spectra. The figures on the left show the interface to identify lines. The observed arc spectra is shown in red and the tabulated lamp emission lines in black. The right-side figure shows how lines are fitted using Lorentzian profiles. The blue lines show the observed arc spectra and the black lines the Lorentzian fits, with the line centre.*

Some of the spectra, including WASP-18's and three of the comparison stars', fell across gaps between chips in the detector. In those cases the ACCESS pipeline extracts the spectra in each detector separately and also independently computes their wavelength calibration fits. In those cases, the final extracted spectra have a small gap which is about 80Å wide in wavelength, as shown in figure 7. The ACCESS pipeline software is smart enough to identify these gaps and match the 2 portions of spectra for each object, to produce a final extracted spectra. An example of those final extracted spectra is shown in figure 7 for WASP-18 and ob004 which corresponds to the brightest comparison star in the field of view.



(a) Spectrum of WASP-18



(b) Spectrum of the brightest comparison star, ob004

Figure 7: *Spectra for a single science image of the first dataset, showing the chip crossover regions with zero counts and absorption lines due to Earth's atmosphere as troughs.*

4 Generation of Light Curves & Detrending

The output of the pipeline described in section 3 contains the extracted, wavelength calibrated spectra for WASP-18b and the five comparison stars for each observation. Additional information in the output includes the exposure time and mid-exposure barycentric Modified Julian Date (BJD−2400000.5 days) of each observation, airmass, and the sky emission level of each observation. These parameters can be used to detrend the light curves and to trace potential systematics in the data, as described below.

I first generated white light curves that were used to obtain the transit parameters such as the centre of transit time, period, inclination, semi-major axis in stellar radii and duration of the transit. These values were then used for transit fitting of the binned light curves to cut down the parameter space used to search for the best fit. This is explained in more detail in section 5. I generated the white light curve of each transit by simply summing the counts of each object, i.e WASP-18 and each comparison star, over that entire wavelength range covered by the IMACS spectra (4,000-10,600Å). This is equivalent to obtaining a photometric transit over a very wide wavelength band. The white light curve of WASP-18 and the brightest comparison star (ob004) are shown in the top diagram of figure 9. Initially I worked with all five comparison light curves, but as explained later, I ended up only using ob004. As seen in figure 9(a), the light curves of each object are plagued by systematic effects due mainly to changes in the Earth's atmosphere. These changes can be mostly compensated for by comparing the light curve of WASP-18 with the light curves of the comparison stars, although some (mainly instrumental) systematic noise will still remain. In the remaining of this section I will detail the different approaches I undertook to model out and remove these systematics; this process is called detrending. I tested the different detrending approaches on the white light curves first, before moving on to the binned light curves.

While detrending the data, dataset 2 was found to still be very dispersive and the start of dataset 3 before the transit couldn't be improved enough to measure the baseline well. These two nights were therefore excluded and I only continued with the rest of the work with the first night of data.

4.1 Systematic Noise

The main difficulty in transmission spectroscopy from the ground is dealing with systematic effects in the data. These include, in order of probable importance:

- Air mass
  - The amount of air the light from the stars has to go through to the telescope, which changes over the night as the targets rise/set.
  - At higher airmass the light from the stars suffer higher extinction, which is also colour dependent so can have a different effect for the target and comparison stars if they have different spectral types (i.e. colours).
- Movement of spectra over time

  - The spectra move over time and fall over different pixels on the detector which have different sensitivities.

- Atmospheric seeing
  - Thermal turbulence in the Earth's atmosphere causes the spectra to have a larger or smaller size on the chip.
  - Need different extraction borders for each image.

- Sky background
  - Can change over the night, e.g. due to the position of the moon with respect to the field being observed.

- Star colour differences
  - As mentioned before in the description of airmass, there are long term smooth trends in each stars' light curves which are due to the target and comparison stars having different spectral types and therefore colours. Different colours suffer different amounts of diffraction by the Earth's atmosphere (especially at high airmass). This is why parts of the light curves of the target and comparisons with high airmass differ more from each other than regions with lower airmass.

- Others
  - Many sources of systematics are not easy to account for, such as the temperature of the detector and small temporal variations in the absorption lines in the Earth's atmosphere (i.e. Telluric lines).

All these effects contribute red noise to the observations. The red noise, however, is correlated to certain conditions as mentioned above and most of it can be modelled out, however, there is always some red noise left in the data because we can never perfectly know all the causes. Typically, it is possible to reduce red noise levels in the optical to $\sim$10% of the original noise level [25]. There is also, as always, random white noise (square root of the number of counts) which cannot be removed from the data. White noise is negligible for the white light curve, but becomes an issue when binning due to less counts, especially at shorter wavelengths (towards 4000Å). This is due to the blaze function of the instrument and can be seen in figure 7 as the shape of the spectra. At the short wavelength end of the spectra, there is zero or few counts which means the light curves in this region are more dominated by both red and white noise, which limits the usable wavelength range.

4.2 Sys-Rem Algorithm

The systematics were first attempted to be removed using my own implementation of the Sys-Rem algorithm, proposed by Tamuz et. al (2005) [26]. This code searches for and

removes linear trends in all light curves, i.e. WASP-18 and the comparison stars. It has been used for data analysis in many large exoplanet searching surveys, such as Super-WASP [27] which found WASP-18b.

The algorithm begins by searching for the best c_i which minimises the expression below,

$$S_i^2 = \sum_j^N \frac{(r_{ij} - c_i a_j)^2}{\sigma_{ij}^2} \tag{4}$$

where r_{ij} is the residual of each measurement - the average flux of the i-th star subtracted from the j-th image taken at airmass a_j, c_i is the effective extinction coefficient of the star and σ_{ij} is the uncertainty in r_{ij}. Although it starts with the airmass at the time of the measurement, the end a_j is not physical and is just the best fit to the trends of the data along with c_i. i ranges from 1 to N and j ranges from 1 to M, where N is the number of stars and M is the number of images from observations. It then uses the best c_i found to minimise equation 5 and find the best a_j and iterates in this way. Once it has found the best c_i and a_j after many iterations, the new residual is made simply by removing the product $c_i a_j$ from each residual as seen in equation 6. The corrected flux is obtained by simply adding the new residual to the stars average calculated at the beginning. The algorithm runs over all is and js and computes their new fluxes, i.e. corrected light curves. The source code of my implementation of the Sys-Rem algorithm is provided in appendix A.

$$S_j^2 = \sum_i^M \frac{(r_{ij} - c_i a_j)^2}{\sigma_{ij}^2} \tag{5}$$

$$r_{ij,new} = r_{ij} - c_i a_j \tag{6}$$

This method had a large drawback however, where the algorithm removed part of the transit from the target and put it into all the comparison star light curves. This was because this is the solution that minimises equations 4 and 5. Obviously this was a problem because the goal is to accurately measure the transit depth, so the detrending method must not change it. A number of ways were tested to attempt to resolve this issue which had varied success. I tried running the algorithm for just comparison stars which should give a good a_j and then found the best c_i for WASP-18b, using the the one which gave the best dispersion in the light curve. This is similar to principle component analysis (PCA) where one looks for trends in only the comparison stars and then apply corrections to all stars, including the target with the transit. For this method to work well you need a good amount of comparison stars to remove the systematics and not just noise associated to one star. This caused a problem when detrending the binned light curves because ob035 (4900-9400Å) and ob018 (5500-10500Å) had a reduced spectral range and there were 4 chip cross overs (one for each of the other stars), all in different places on the detector with the region around the cross overs also giving poor quality data. I also tried to correct the effect by putting the 'inverse transit' back into the target's light curve and removing it from the comparison light curves. I tried fitting functions to the inverse transit in the comparison stars and correcting

all light curves. I also tried computing the deviation of each flux measurement in the inverse transit from the average of the out-of-transit portions for the comparison light curves; then using the average deviation from the comparison stars to correct the transit light curve. However, these methods were insufficient to correct for the effect without altering the light curves away from their real shape, i.e. it was not possible to perfectly correct the transit and therefore not a viable method to use.

Using each comparison light curve many times for Sys-Rem reduced the effect, with it becoming unnoticeable when the comparison stars were used more than 30 times each; this can be seen in top of figure 8. However, as can be seen in the bottom panel of figure 8, this solution had the adverse effect of increasing the dispersion of the target light curve to almost double when the inverse transit becomes unnoticeable. This effect therefore gets overlooked in a large exoplanet transit survey where they have many thousands of stars in the field of view, but may increase the dispersion of their light curves sufficiently for this algorithm to not be ideal method to use. Sys-Rem is not well suited for datasets in which only a handful of comparison stars are available.

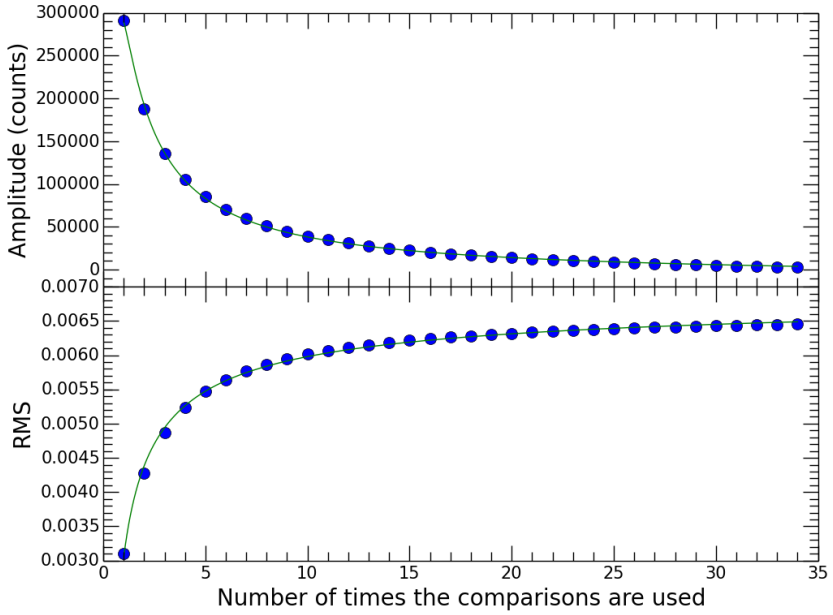


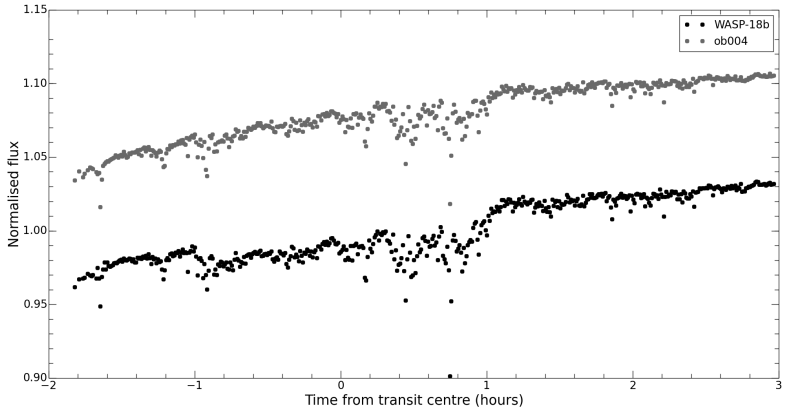
Figure 8: *Amplitude of the 'inverse transit' (top) and dispersion of the transit light curve (bottom) against the number of times the comparison stars were each used.*

4.3 Polynomial Method

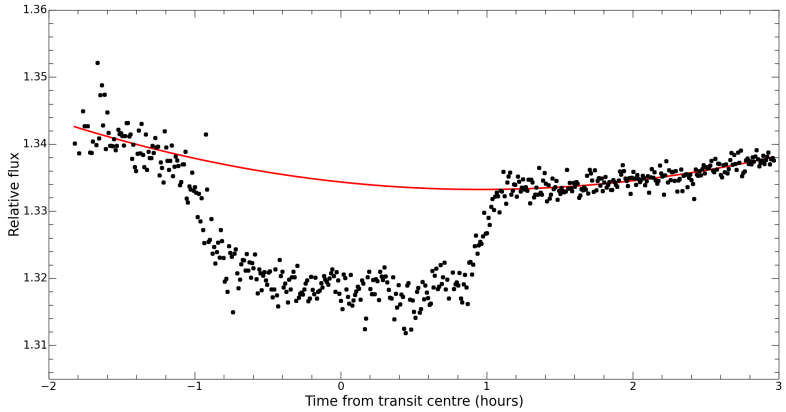
The next technique I used was the polynomial method used for the same work but on the super-Earth GJ 1214b by Bean et al. (2010) [28]. The method for this was to simply

divide the flux of the target by the flux of the brightest comparison (ob004 in my case) and then fit a polynomial to the out-of-transit part of the light curve to model the long term trend. The long term trend (see figure 9(b)) is most likely due to the slight colour difference between ob004 and WASP-18, therefore causing light from them to behave differently in the Earth's atmosphere; i.e. bluer stars being refracted more than redder ones. The trend is well modelled over time as a second order polynomial. The effect was close to linear with air mass but the polynomial fit with time was superior. I used only the brightest comparison because it behaved the most similar to WASP-18 and had the best dispersion of the 5 comparison stars. I tried using the sum of the light from all comparison stars, but like Bean et al., I found this introduces more noise due the other 4 comparison stars being much dimmer and therefore their light curves being more dispersive. This was evident in the out-of-transit residuals to the transit fit. This method was superior to the Sys-Rem algorithm because by using only 2 stars, there were only 2 chip cross overs to deal with and apart from the crossings the whole spectral range was usable. The python code I wrote to carry out this detrending method is provided in appendix B.

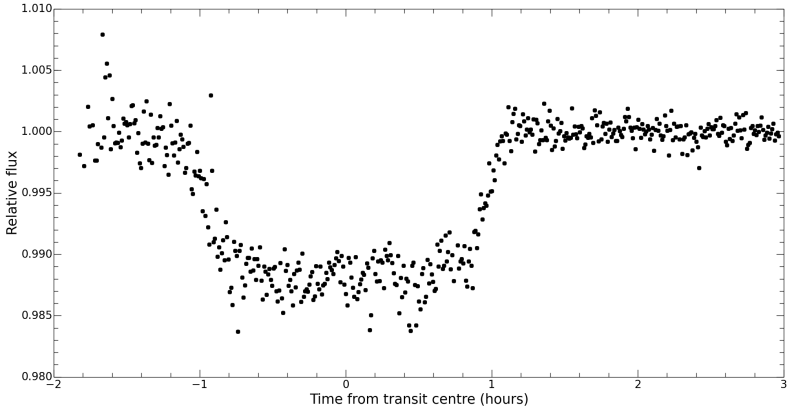
Figure 9 shows the steps of the polynomial method for the white light curves, i.e. all the wavelength bins summed. Part 9(a) shows the data after it has been reduced by the pipeline for the target (black) and the brightest comparison star (grey). The common trends between the two light curves can be seen, which are mostly removed by the dividing step. Part 9(b) shows the second order polynomial fit to the out-of-transit portions of the light curve after dividing and 9(c) shows the corrected white light curve after both amendments with 2 outlying data points clipped which were outside 3σ (3 standard deviations discrepant).



(a) *Reduced, but undetrended white light curves for WASP-18b and the brightest comparison*



(b) *Second order polynomial fit to the light curve obtained after dividing the target by the comparison*



(c) *Final corrected white light curve*

Figure 9: *Steps of the polynomial method*

I then tried to improve the dispersion by looking for trends in the light curve with some of the systematics mentioned before, namely the seeing and the sky background variations over the night, similar to work done by Wilson et. al (2015) [29]. A value of the seeing was obtained at each time of measurement by taking the raw files from the observations and computing the full-width-half-maximum (FWHM) of a Gaussian-like cut across the target's spectrum for each image. To do this, I took the files for the chip containing most of the spectrum of WASP-18, fixed a row of pixels in the spectral direction free of Earth's absorption lines and then for each image, fitted a Gaussian function along the row ± 15 pixels of the rough spectrum centre, which can be seen in figure 10. I then calculated the FWHM of this Gaussian by finding the difference in the pixel numbers which were half the peak of the Gaussian, allowing for fractions of a pixel number for more accuracy. A plot of the seeing varies over the night is shown in figure 11 along with a plot of FWHM vs residual. The code used for this can be seen in appendix C.

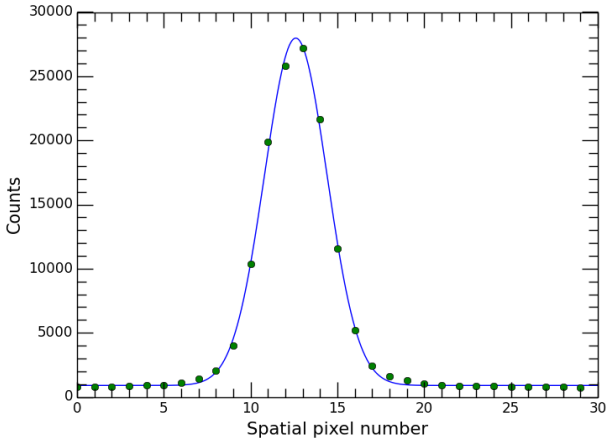
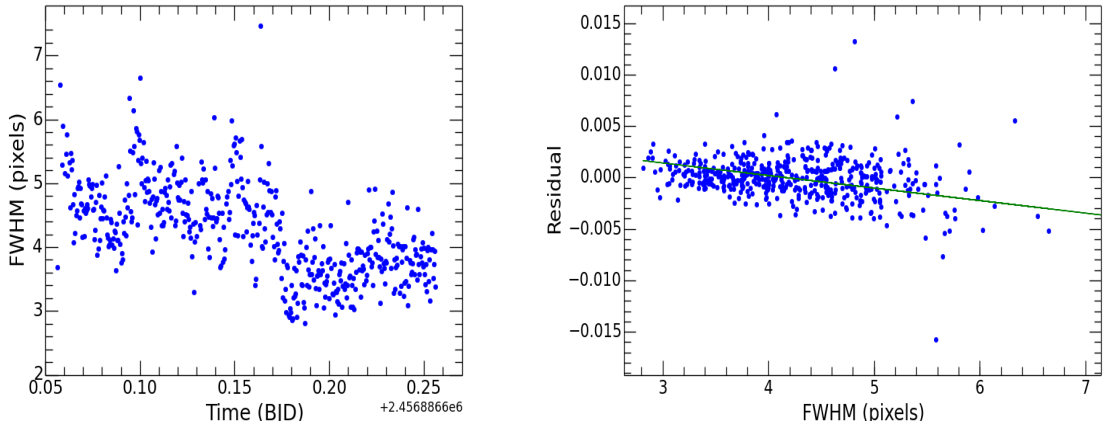


Figure 10: *A cut across the spectrum of WASP-18 with a Gaussian function fitted against pixel in the spatial direction.*

A measurement of the sky emission per pixel in the dispersion direction was obtained at the data reduction stage using the pipeline, as mentioned before. The average value of the sky emission was plotted against the residuals for the white light curve, i.e. the data subtracted by the transit model fit, shown in figure 12. The effect of the movement of the spectra over time could not be evaluated, given the difficulty in measuring the change in detector position accurately.

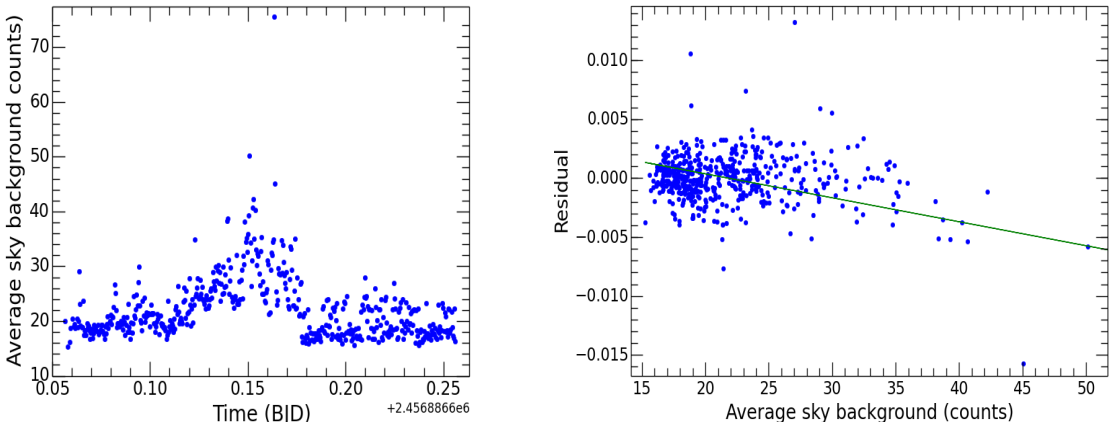
Much like Bean et al., I didn't see any strong correlations after reference star and long term trend corrections to the seeing or sky emission level. This hints that these systematics are fully accounted for in the reference star correction; i.e. these stars spectra have a very similar seeing and sky emission response. Also, the seeing variations may not have had much of an effect on the data due to the large extraction apertures used (± 15 pixels), which meant no star light was lost during extraction up to a seeing of 3.5 arcsec; i.e. perfectly fine for



(a) How the seeing varied over the night

(b) Seeing against residual

Figure 11: *The effect of seeing on the data.*



(a) How the sky emission level varied over the night

(b) Sky background against residual

Figure 12: *The effect of sky background on the data.*

dataset 1. After applying this polynomial fit detrending method, the white light curve of WASP-18b was ready for transit fitting, as described in the next section.

5 Transit Fitting

The detrended white transit light curve was fitted using an IDL (Interactive Data Language) package called EXOFAST [30], explained below. The inputs to this program were a file containing a column for the detrended flux, uncertainty in the flux & each time of measurement in barycentric Julian Date, and a list of starting values for some of the system parameters,

called priors. These were obtained from the literature for the white light curve fitting, using the WASP-18b discovery paper [19] and Triaud et. al 2010 [31]. I used priors for the time of centre transit, period, stellar surface gravity, stellar effective temperature and stellar metallicity. Table 5 shows the values used for the white light curve. Priors were used to greatly cut down the region of parameter space used to search for the best fit values. This allows faster convergence and more accurate results due to finding the global minimum in a relatively well-behaved region of parameter space. With good priors for the period and time of the transit centre, it is possible to get good estimates for the other parameters. I fixed the planet's orbit to be circular due to the very low eccentricity found in the literature (0.00848±0.00095, Triaud 2010). I also used the iterative sigma-clipping process which cut any points (typically 2-15) more than 3σ (3 standard deviations) from the transit fit, and then re-ran the fitting and clipping until no more points were cut. This minimises the dispersion of the fit due to outliers in the data.

Prior	Value	Uncertainty
Time of centre transit (BJD)	2454664.90531	±0.00017
Period (days)	0.94145290	$\pm 8.6 \times 10^{-7}$
Stellar surface gravity (log g)	4.367	±0.096
Stellar effective temperature (K)	6400	±100
Stellar metallicity	0.00	±0.09

Table 5: *Table of priors used for the white light curve transit fit*

EXOFAST also fits for stellar parameters: mass, radius, luminosity, density, planetary parameters: semi-major axis, radius, equilibrium temperature, incident flux on the planet and transit parameters: quadratic limb-darkening coefficients, baseline flux of the observations, time of eclipse, inclination, impact parameter, transit depth, FWHM duration, ingress/egress duration and total transit duration.

EXOFAST begins by fitting the transit model to the data. It finds the best model to be the one which minimises equation 7, for dataset D with uncertainties σ, generated model M and i corresponding to each of the n data points in the dataset. The transit model is non-linear and therefore the best fit parameters which minimise χ^2 must be determined numerically. There are no generic algorithms for a global parameter space which is why the parameter space searched is restricted to a region close to the global minimum by using priors; many routines can robustly find the global minimum from this region. EXOFAST uses its own implementation of the AMOEBA routine (Nelder & Mead 1965 [32]) which given a starting point and stepping scale, will work through the parameter space to find the minimum using the χ^2 at each step to determine the next step.

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{D_i - M_i}{\sigma_i} \right)^2 \tag{7}$$

If the star were uniformly bright, the relative flux observed during transit would be only a function of the transit geometry, i.e. the distance from centre of the planet to the centre

of the star, the planetary & stellar radii and time. However, stars are not uniformly bright and the flux received drops towards the limbs of the star (limb darkening effect). This effect is wavelength dependent and has a larger effect towards bluer wavelengths. For main sequence stars, like WASP-18, limb darkening is well modelled by functions of $\mu = \cos(\theta)$ where θ is the angle between the line of sight and radiation from the star. The value of μ is 1 at the centre of the star and 0 at the limbs. EXOFAST uses a quadratic limb darkening law (equation 8) over the linear law for higher precision. The law is sufficient to describe transit light curves with a precision of $10^{-4}(\frac{R_p}{0.1R_s})^2$ [33], which is better than can currently be achieved from the ground [34]. I used an option in the program to select the band (V) for the data, so EXOFAST could obtain quadratic limb darkening coefficients from tables of Claret & Bloemen (2011) [35].

$$\frac{I(\mu)}{I(1)} = 1 - u_1(1-\mu) - u_2(1-\mu)^2. \tag{8}$$

To estimate the uncertainties of each parameter, EXOFAST uses a variation of Markov Chain Monte Carlo (MCMC) called Differential Evolution MCMC (DE-MC) in which multiple chains are ran in parallel, allowing the chains to learn from each other. The EXOFAST MCMC best fit is shown in figure 13 and the values of each parameter for the fit and their corresponding uncertainties are summarised in table 6. In addition, and as example of the MCMC results, I show the MCMC distributions for some of the main parameters of the fit yielded by the IDL code in figure 14. The distributions shown have only 10% of the actual steps used. This is because EXOFAST reduces the size of the output IDL save files by only saving 1 in 10 steps. The best fit and uncertainty calculations use the full chains however. A total of 5.7% of the iterations were discarded before the 'burn-in', to eliminate biases due to the starting conditions. EXOFAST computes 1σ uncertainties for each of the output parameters using it's MCMC, which is easy to do because the distributions resemble Gaussian functions. The final parameter value is taken to be the median of the distribution and the 34% confidence interval either side is taken to be the uncertainty.

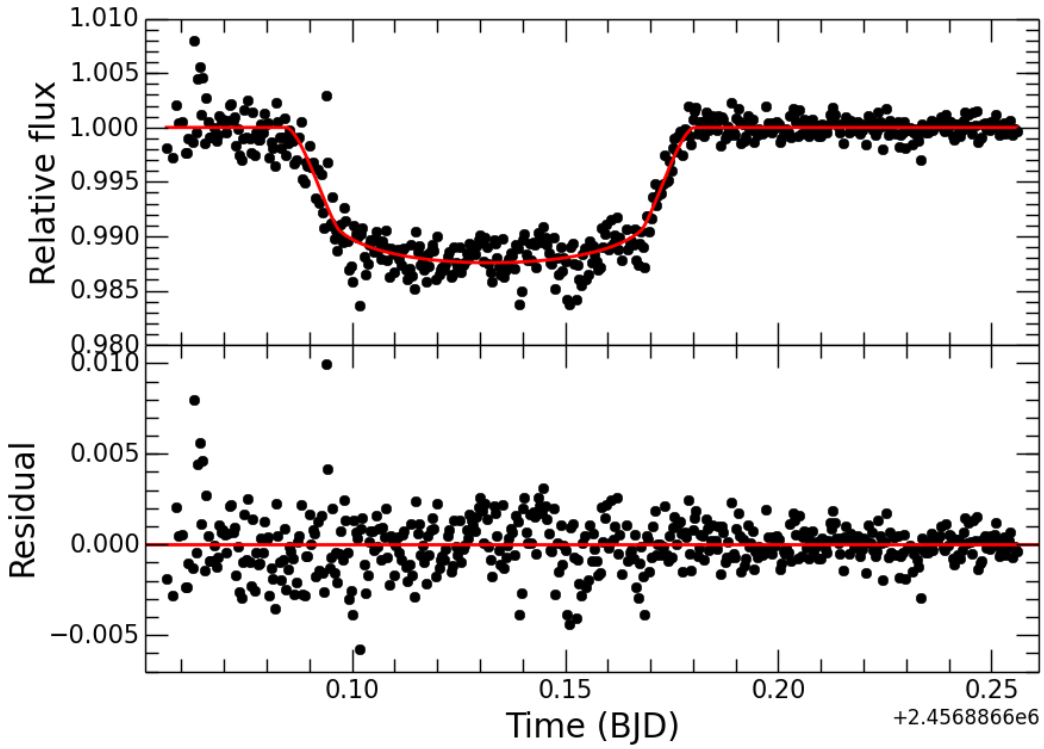


Figure 13: *The EXOFAST best fit to the detrended white light curve.*

Parameter	Value	Uncertainty
Stellar parameters:		
Mass ($M_\odot$)	1.234	$\pm$0.066
Radius ($R_\odot$)	1.197	$\pm$0.063
Luminosity ($L_\odot$)	2.18	$\pm$0.30
Surface gravity (log g)	4.373	$\pm$0.038
Effective temperature (K)	6413	$\pm$99
Metallicity	-0.006	$\pm$0.089
Planetary parameters:		
Period (days)	0.94145290	$\pm 8.4 \times 10^{-7}$
Semi-major axis (AU)	0.02016	$\pm$0.00036
Radius (R_J)	1.228	$\pm$0.077
Equilibrium temperature (K)	2382	$\pm$67
Incident flux (10^9 erg s^{-1} cm^{-2})	7.31	$\pm$0.83
Transit parameters:		
Time of centre transit (JD)	2456887.67401	$\pm$0.00033
Radius of planet in stellar radii	0.1054	$\pm$0.0015
Semi-major axis in stellar radii	3.62	$\pm$0.17
Linear limb-darkening coeff, u_1	0.316	$\pm$0.047
Quadratic limb-darkening coeff, u_2	0.290	$\pm$0.049
Inclination (degrees)	82.8	$\pm$1.4
Impact Parameter	0.452	$\pm$0.070
Transit depth	0.01111	$\pm$0.00032
FWHM duration (days)	0.0751	$\pm$0. 0012
Ingress/egress duration (days)	0.01019	$\pm$0.00094
Total duration (days)	0.0853	$\pm$0.0021

Table 6: *Output parameters with their uncertainties from the EXOFAST MCMC.*

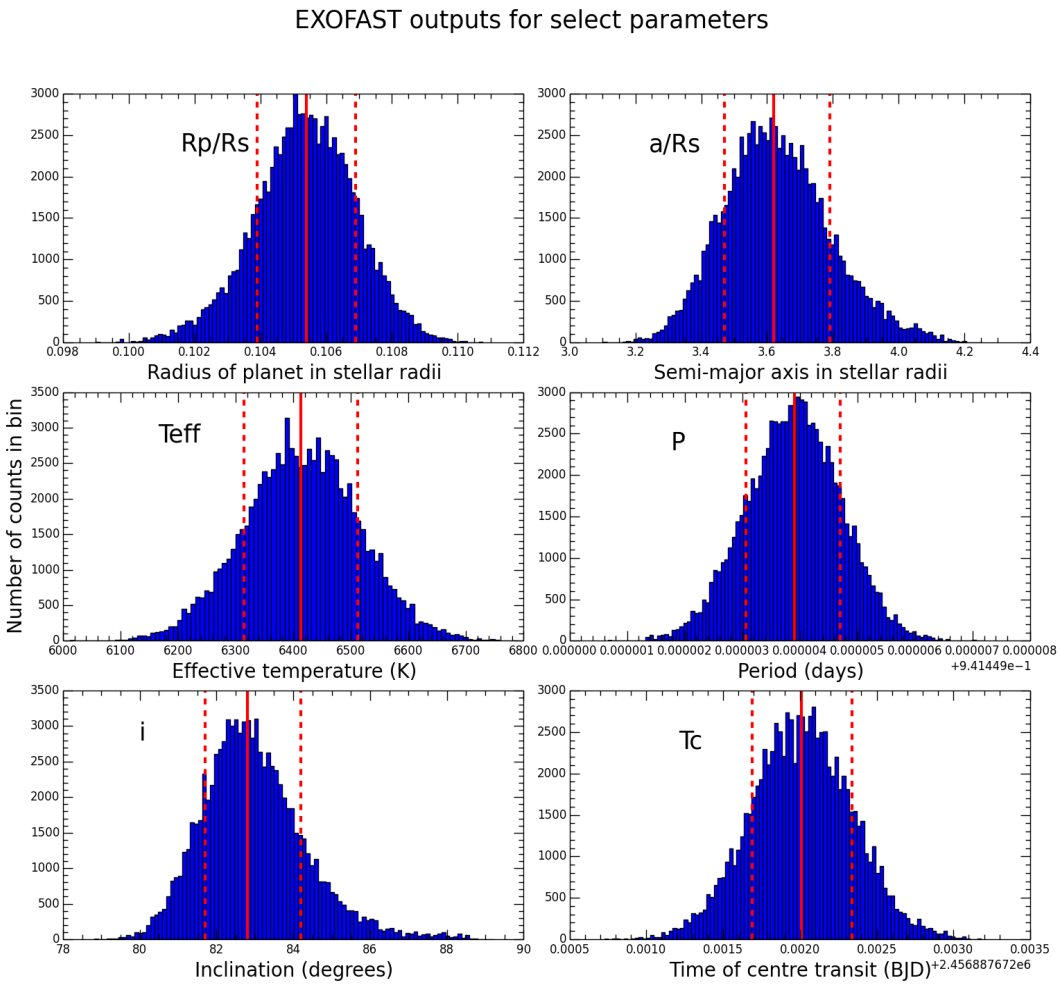


Figure 14: *Distributions from the EXOFAST MCMC with the data binned into 100 bins according to the parameter value. The solid red line shows the value of the best fit and the two dotted lines show the 1σ deviations from the best fit.*

6 Binning

The white light transit light curve fit presented in the previous section provides the average parameters over the whole wavelength range, in particular the average atmospheric radius of the planet over the entire visible band (4000-10600Å). The main objective of this work, however, is to search for variations in transit depth (R_p/R_*) as a function of wavelength, since these variations provide information about the chemical composition of the planets

atmosphere[1]. In this section, I explain the work done to obtain R_p/R_* vs wavelength over the entire visible wavelength band for WASP-18b. The first step of this process was to bin the spectra of WASP-18b in small wavelength intervals.

Binned light curves were obtained simply by summing the counts of the spectra in a defined wavelength region, e.g. 5000-5200Å. The data were binned into 38 bins of 200 angstroms for most of the spectra to get enough photons for good signal to noise, but using 100 angstrom bins around the theoretical spectral features of sodium and potassium (see figure 3) to get better wavelength resolution in those regions. I used 7 100Å bins for the Na line and 3 for the K line. There were also 3 bins of different wavelength ranges over parts of the spectrum that were unusable: the first bin with very few counts and the two chip crossings. The detrending and transit fitting analysis of each of the 38 bins analogous to the ones explained for the transit white light curve in sections 4 and 5. Each bin was detrended separately using the polynomial method explained in section 4.3 and fitted with EXOFAST. For the binned light curves, I used the output values from the white light curves as priors for the individual bins. These values are shown in table 7 and are very close to the priors used for the white light curve itself. I also added a prior for the orbital inclination, i, from the EXOFAST output of the white light curve. It was necessary to add this prior, because for some bins EXOFAST would prefer a very large radius ($>15R_J$) with a lower orbital inclination angle ($<65°$).

Table 8 shows the bins used and the planet-to-star radius ratios obtained for each of them. The uncertainties given are the 68% confidence intervals from the EXOFAST MCMC. Figures 15, 16 and 17 show the spectroscopic light curves obtained for each bin in table 8 with their best-fitting transit models and residuals to these fits. The light curves are split into 3 figures for clarity and each light curve & set of residuals are offset slightly by flux to make the bins easier to distinguish. The best data occurs in the wavelength regions with the most counts (see figure 7), with more disperse data at the edges of the spectrum.

Prior	Value	Uncertainty
Time of centre transit (JD)	2456887.67401	±0.00033
Period (days)	0.94145290	$\pm 8.4 \times 10^{-7}$
Stellar surface gravity (log g)	4.373	±0.038
Stellar effective temperature (K)	6413	±99
Stellar metallicity	-0.006	±0.089
Inclination (degrees)	82.8	±1.4

Table 7: *Table of priors used for the binned light curves*

[1]Searching for variations of R_p/R_* vs. wavelength is the overall goal of the ACCESS project, but for this target in particular, the goal was to establish with which precision we can measure these variations by choosing a planet expected to have very shallow atmospheric features and therefore a flat R_p/R_* vs. λ curve.

Bin number	$\lambda_{\rm min}$ (Å)	$\lambda_{\rm max}$ (Å)	λ range (Å)	R_p/R_*
0	4000	4143	143	-
1	4143	4343	200	-
2	4343	4543	200	-
3	4543	4743	200	$0.1078\ ^{+0.0025}_{-0.0025}$
4	4743	4943	200	$0.1083\ ^{+0.0023}_{-0.0022}$
5	4943	5143	200	$0.1069\ ^{+0.0020}_{-0.0020}$
6	5143	5343	200	$0.1108\ ^{+0.0017}_{-0.0017}$
7	5343	5543	200	$0.1096\ ^{+0.0019}_{-0.0018}$
8	5543	5643	100	$0.1121\ ^{+0.0031}_{-0.0031}$
9	5643	5743	100	-
10	5743	5843	100	$0.1128\ ^{+0.0030}_{-0.0031}$
11	5843	5943	100	$0.1120\ ^{+0.0028}_{-0.0028}$
12	5943	6043	100	$0.1131\ ^{+0.0029}_{-0.0031}$
13	6043	6143	100	$0.1030\ ^{+0.0024}_{-0.0024}$
14	6143	6243	100	$0.1127\ ^{+0.0026}_{-0.0027}$
15	6243	6443	200	$0.1127\ ^{+0.0035}_{-0.0036}$
16	6443	6643	200	$0.1132\ ^{+0.0015}_{-0.0016}$
17	6643	6843	200	$0.1130\ ^{+0.0055}_{-0.0055}$
18	6843	7132	289	-
19	7132	7332	200	-
20	7332	7532	200	$0.1044\ ^{+0.0021}_{-0.0022}$
21	7532	7632	100	$0.1072\ ^{+0.0044}_{-0.0044}$
22	7632	7732	100	$0.1082\ ^{+0.0045}_{-0.0047}$
23	7732	7832	100	-
24	7832	8000	168	-
25	8000	8200	200	$0.1150\ ^{+0.0046}_{-0.0045}$
26	8200	8400	200	$0.1122\ ^{+0.0043}_{-0.0046}$
27	8400	8600	200	$0.1105\ ^{+0.0019}_{-0.0020}$
28	8600	8800	200	$0.1146\ ^{+0.0068}_{-0.0070}$
29	8800	9000	200	$0.1067\ ^{+0.0042}_{-0.0044}$
30	9000	9200	200	$0.1083\ ^{+0.0026}_{-0.0026}$
31	9200	9400	200	$0.1051\ ^{+0.0040}_{-0.0041}$
32	9400	9600	200	$0.1039\ ^{+0.0032}_{-0.0032}$
33	9600	9800	200	$0.1115\ ^{+0.0046}_{-0.0047}$
34	9800	10000	200	$0.1147\ ^{+0.0040}_{-0.0040}$
35	10000	10200	200	$0.1094\ ^{+0.0032}_{-0.0031}$
36	10200	10400	200	$0.1073\ ^{+0.0016}_{-0.0016}$
37	10400	10600	200	$0.1129\ ^{+0.0056}_{-0.0055}$

Table 8: *Table of bins used*

The resulting transmission spectrum, i.e. R_p/R_* against λ, of WASP-18b is shown in figure 19. A total of 8 bins were not used in the plot. The first 3 bins (4000-4743Å) didn't have enough counts and were dominated by noise, limiting the performance of EXOFAST. For bin 9 (5643-5743Å), the target and brightest comparison behaved very differently at the start of observations which continued in to the transit. I could not identify the source behind this. Bins 18, 19, 23 and 24 (6843-7332Å & 7732-8000Å) were poor due to the spectra of either the target or the comparison ob004 crossing detector or falling close to the chip edges.

The R_p/R_* values in figure 19 and table 8 have also been corrected for a trend in EXOFAST with the transit ingress/egress duration, τ. After fitting models to the transits with EXOFAST, it was discovered that the best fits all used different values of τ. Thinking physically, τ should be the same for any bin or wavelength, but I found a direct correlation between τ and transit depth obtained using EXOFAST. I corrected for this by fitting a straight line to τ minus the average τ against R_p/R_* obtained for each bin, which can be seen in figure 18. After this correction, no other such correlations were found with any parameters.

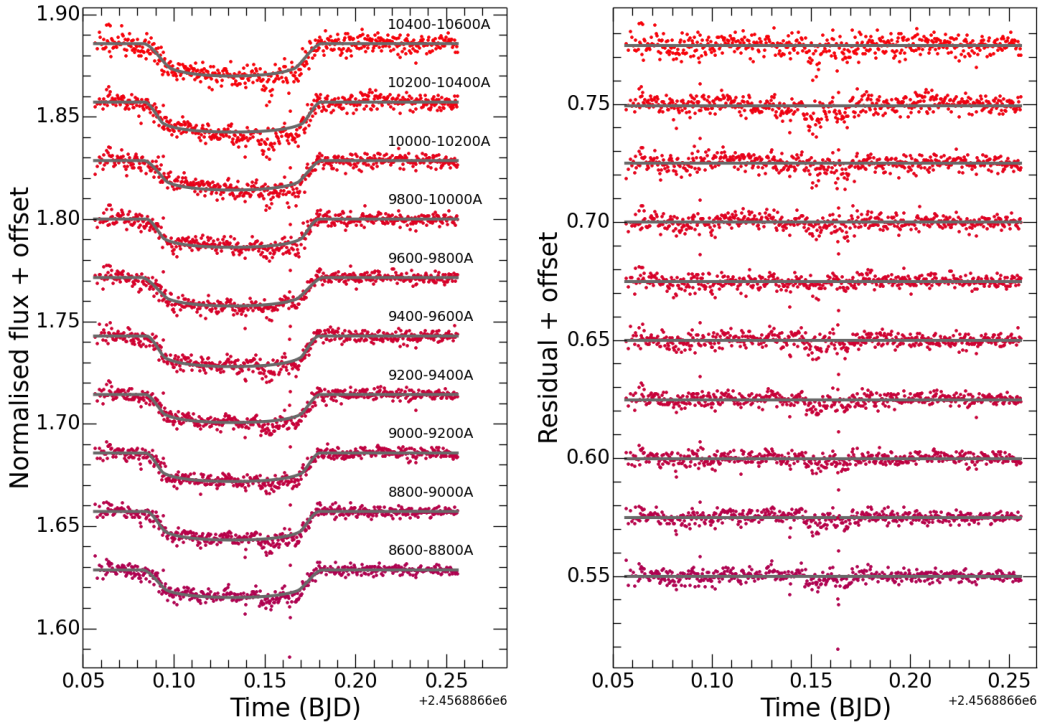


Figure 15: *Left: detrended binned light curves with the best fit transit model for bins between 8600Å and 10600Å. Right: residuals to the fit with the zero line plotted through them.*

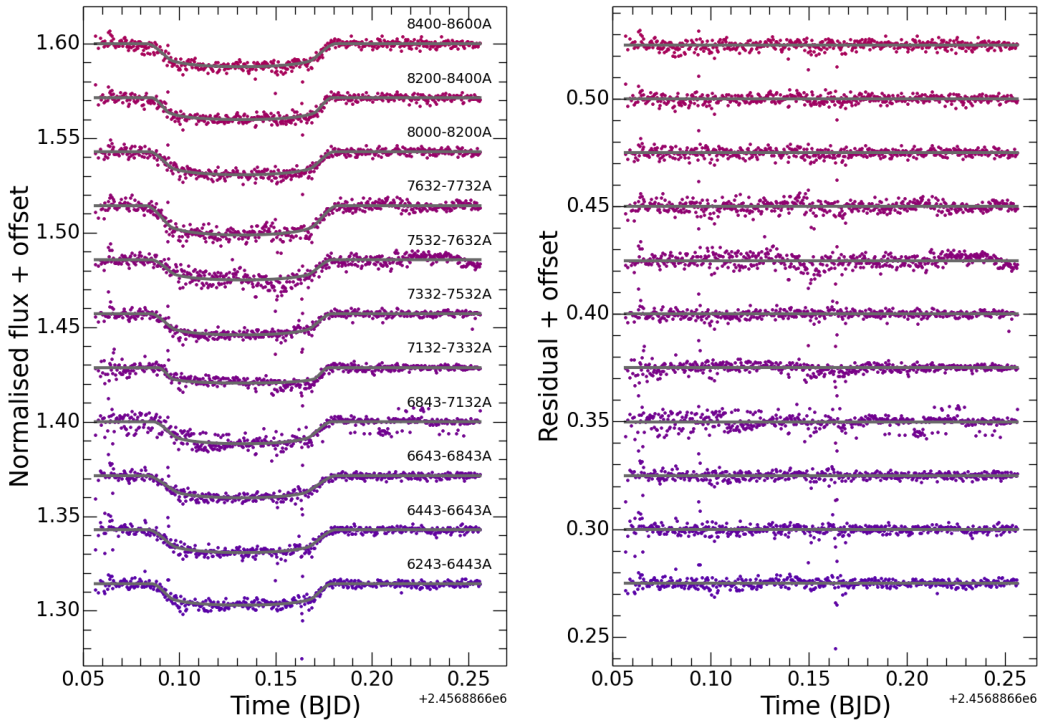


Figure 16: *Left: detrended binned light curves with the best fit transit model for bins between 6243Å and 8600Å. Right: residuals to the fit with the zero line plotted through them.*

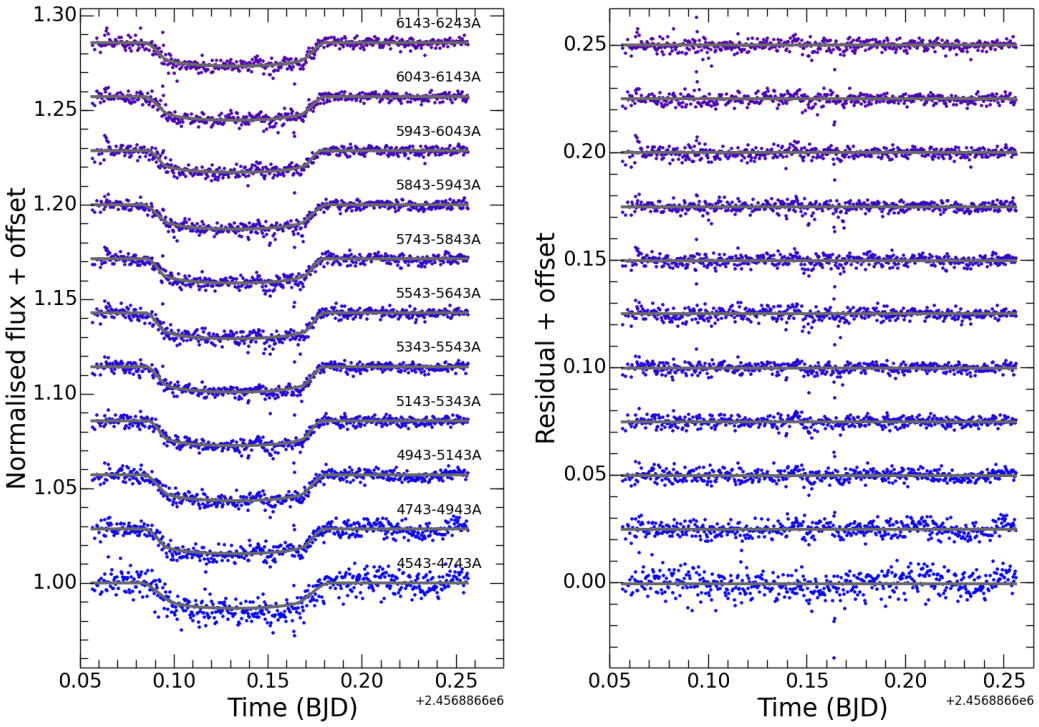


Figure 17: *Left: detrended binned light curves with the best fit transit model for bins between 4543Å and 6243Å. Right: residuals to the fit with the zero line plotted through them.*

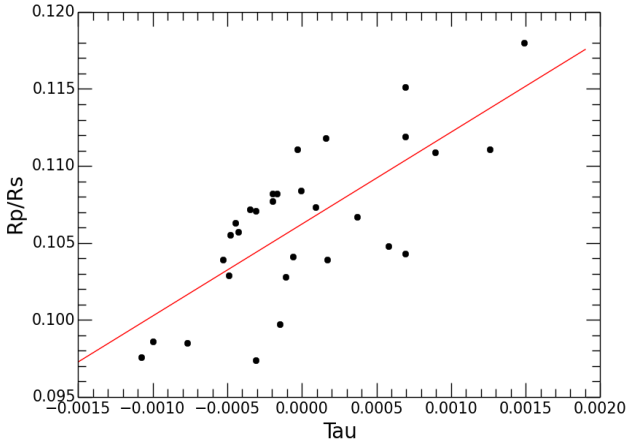


Figure 18: *The ingress/egress time for each bin against R_p/R_* from the MCMC best fit.*

7 Results

When comparing the theoretical models to the R_p/R_* spectrum of the data after binning (see figure 19) I found that the variations of the models were much less than the error bars of my data points. This is due to the scale height of the atmosphere of WASP-18b. The uncertainties I find are actually very similar to those attained by ACCESS in other work on WASP-31b but the scale heights of the planets are very different. WASP-31b is also a hot-Jupiter, with a radius larger than WASP-18b and a much smaller mass. The parameters of the two planets are compared in table 9, with the WASP-31b information taken from it's discovery paper [36]. Equation 3 is repeated below along with a table to show why and by how much these atmospheres differ. I assumed that both planets have the same mean molecular mass and computed the ratio of scale heights, as seen in equation 9, by simply comparing the planets mean atmospheric temperature and optical surface gravity (i.e. their masses and radii), which gave a rough answer of 0.04 or 1/25. Variations in the models are proportional to the scale height, so therefore spectral features are 25 times easier to see in WASP-31b compared to WASP-18b. A comparison of the models for WASP-18b and WASP-31b can be seen in figure 19, showing a large difference in the size of expected spectral features. The full lines show models for WASP-18b and the dotted lines are models for WASP-31b.

As explained before, only one of the datasets for WASP-18b was useful, therefore the errorbars in the planet's spectrum in figure 19 could have been $\sqrt{3}$ times smaller if I had three good sets of binned transit light curves. In addition, the dispersion in the final spectrum (figure 19) could have been further improved by averaging the results of the three datasets. The extra datasets would have also served to trace potential systematics in derived exoplanet spectrum. With precision obtained from combining three datasets, I would have been able to detect Na and K features of a planet like WASP-31b, if it had a clear atmosphere.

This strongly shows the need for 'good' scale height targets, i.e. targets that will have detectable spectral features. This may even be the top priority when choosing which exoplanets to study the atmospheres of when using transmission spectroscopy. An estimate for the scale height of an exoplanet needs values for the planet's mass, radius and temperature. These can be obtained from the observations of the planets detection and a follow up observation, using both the transit and radial velocity methods.

$$h = \frac{k_B T}{m_H \mu_m g} \tag{3}$$

$$\frac{h_{18}}{h_{31}} = \frac{T_{18} g_{31}}{T_{31} g_{18}} = \frac{T_{18} M_{31} R_{18}^2}{T_{31} M_{18} R_{31}^2} = 0.04 \tag{9}$$

Around the sodium spectral feature (the doublet line at 5893Å), the errorbars in the final spectrum are typically ± 0.003 in R_p/R_*. The theoretical models for WASP-18b had a ΔR (from the tip of the peak to the lowest point of the feature trough) of roughly 0.0003, i.e. 10 times lower than the uncertainty in the data points. However, the clear atmosphere models for WASP-31b had a ΔR of about 0.0077, which is comparable to the uncertainties

Property	WASP-18b	WASP-31b
Mass	10.30 M_J	0.478 M_J
Radius	1.106 R_J	1.537 R_J
Temperature	2384 K	1575 K

Table 9: *Values used for scale height ratio estimate*

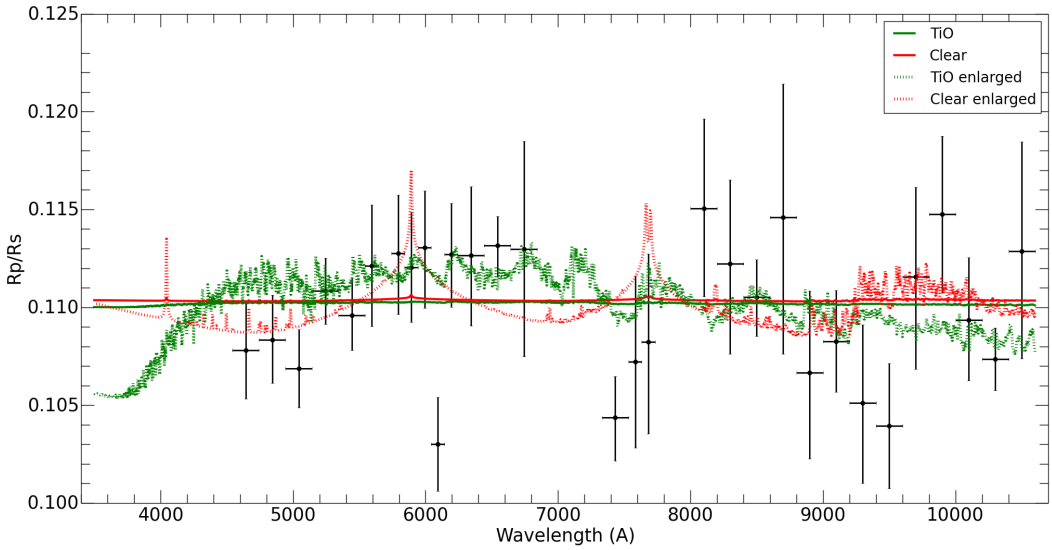


Figure 19: *The final spectrum of WASP-18b, wavelength against planetary radius in stellar radii. The green lines are for an atmosphere containing TiO and the red lines are for a clear atmosphere with Na, K and H_2O. Solid lines are the models at the correct scale for WASP-18b and the dotted lines are rescaled models for a planet with a scale height 25 times larger than WASP-18b (such as WASP-31b).*

in the data. With the 2 extra datasets, i.e. errorbars then typically around ± 0.002, a 3σ detection of the sodium doublet line would be attainable. Similarly for the potassium feature at 7682Å, the errorbars are typically ± 0.0045, with the original models ΔR roughly 0.00024 and the scaled ΔR about 0.006. With 2 extra datasets the uncertainty could be improved to ± 0.003, allowing for just a 2σ detection of potassium.

The spectral feature of H_2O extends over a large wavelength range which means the usage of larger bins to obtain more signal to noise could reduce the errorbars. However, the feature is shallow and is at the very red end of the spectrum which means the data is quite noisy. This makes a significant water detection impossible for data of this quality and a scale height similar to WASP-31b. Titanium oxide is also difficult to measure the wavelength variations of. The blue end of the TiO model gives a good shape to test, but the usable wavelength range of the spectrum doesn't start until this shape has flattened out. This means the only

way to test the TiO model is to measure variations in the peaks and troughs along the spectra. The ΔR for these variations is about 0.0001 for the original model and 0.003 for the scaled one. With errorbars of $\pm$0.003, even with 2 more sets of data these variations would not be possible to detect currently even for scale height comparable to WASP-31b. The best target currently to search for TiO would probably be the bloated hot-Jupiter, WASP-76b [37], with a mass $0.92M_J$, a radius $1.83R_J$ and a temperature 2190K, giving it a scale height 27 times larger than that of WASP-18b. However, the observations would have to be done from space to heavily reduce the noise and therefore uncertainties or otherwise require a minimum of 9 good transits from the ground, assuming the same precision that I achieved.

The scale height is proportional to ΔR and $\Delta R \geq 3\sigma$ is required for a 3σ detection, so from this, an estimate of the needed scale height can be found, i.e. the minimum scale height to detect optical atmospheric features in exoplanets. Using basic algebra, it is easy to see how many times larger the scale height needs to be compared to WASP-18b. From equation 10 (subscript n for the needed value), the scale height needs to be a factor of $3\sigma/\sqrt{3}\Delta R$ larger, assuming data for 3 transits with the same precision as the dataset I used. Table 10 shows the factors larger in scale height required to detect each spectral feature in an exoplanet's atmosphere.

$$\frac{h_n}{h} = \frac{\Delta R_n}{\Delta R} = \frac{3\sigma}{\Delta R} \tag{10}$$

Spectral feature	σ	ΔR	Scale height required (x h_{18})
Na	$\pm$0.003	0.00031	17
K	$\pm$0.0045	0.00024	32
H_2O	$\pm$0.003	0.00010	52
TiO	$\pm$0.003	0.00011	47

Table 10: *Scale heights greater than WASP-18b needed to detect each spectral feature at 3σ.*

8 Discussion & Conclusions

The precision I obtained with one transit proves that the atmospheres of exoplanets with very small scale heights, such as WASP-18b, will be difficult to study unless we have dedicated telescopes that could collect a large number of transits of any given target. However, for planets with larger scale heights, the precision obtained with IMACS is sufficient to measure features like the sodium and potassium absorption, in the case of planets with clear atmospheres. Using the example of WASP-31b from section 7, and using three good transits (i.e. improved errorbars by a factor of $\sqrt{3}$), it is possible to detect Na and K in the atmosphere of that planet at 3σ and 2σ respectively.

The results of my work prove that the atmospheric properties of gas-giant planets can be studied at visible wavelengths, and from the ground, with existing instrumentation using large telescopes. However, these instruments were built for other less precision-demanding

science, which plague our observations with very difficult to track systematics. Ideally, we would like to have future generations of instruments in large telescopes built especially to reduce these systematics at or below the level of exoplanetary atmospheric signals. Such a desired telescope would have a large field view to utilise many comparison stars around the target. It would also have long detector chips for each star's spectrum to fall all on the same chip to avoid gaps in the final spectrum. For detailed studies of exoplanetary atmospheres, targets should be prioritised according to their scale heights, although to do this one needs good estimates of the planet's radius, mass and temperature, which isn't the case for most known exoplanets.

Another way to improve the precision in the data points on the final spectrum, would be to improve the detrending method. Using a more powerful detrending technique, such as PCA, would improve the errorbars by reducing the dispersion in the binned light curves. With enough of an improvement in the detrending method, a 3σ detection of TiO in an exoplanet's atmosphere would be possible from the ground with 3 good datasets of a suitable target, such as WASP-76b.

Acknowledgements

I would like to thank the University of Southampton for giving me the opportunity to be on the year abroad course and the Harvard-Smithsonian Center for Astrophysics for giving me access to all their facilities and the chance to meet and see talks by world leading astronomers on a daily basis. In particular I thank my supervisor, Mercedes López-Morales, for all her knowledge and constant enthusiasm. I thank the ACCESS collaboration for kindly letting me work as part of their team. Lastly, I thank Phil Charles and Jeremy Drake for organising the program and for all their help over the course of my work.

References

[1] Jean Schneider, Cyrill Dedieu, Pierre Le Sidaner, Renaud Savalle, and Ivan Zolotukhin. Defining and cataloging exoplanets: the exoplanet.eu database. *Astronomy & Astrophysics*, 532:A79, 2011.

[2] F. Pepe, A. Cameron, D. Latham, E. Molinari, S. Udry, A. Bonomo, L. Buchhave, D. Charbonneau, R. Cosentino, C. Dressing, et al. An Earth-sized planet with an earth-like density. *Nature*, 503(7476):377–380, 2013.

[3] Z. Berta, D. Charbonneau, J.M. Désert, E. Kempton, P. McCullough, C. Burke, J.J. Fortney, J. Irwin, P. Nutzman, and D. Homeier. The flat transmission spectrum of the super-Earth GJ1214b from Wide Field Camera 3 on the Hubble Space Telescope. *ApJ*, 747(1):35, 2012.

[4] L. Kreidberg, J. Bean, J.M. Désert, B. Benneke, D. Deming, K. Stevenson, S. Seager, Z. Berta-Thompson, A. Seifahrt, and D. Homeier. Clouds in the atmosphere of the super-Earth exoplanet GJ1214b. *Nature*, 505(7481):69–72, 2014.

[5] D.K. Sing, F. Pont, S. Aigrain, D. Charbonneau, J.M. Désert, N. Gibson, R. Gilliland, W. Hayek, G. Henry, H. Knutson, et al. Hubble space telescope transmission spectroscopy of the exoplanet HD 189733b: high-altitude atmospheric haze in the optical and near-ultraviolet with STIS. *MNRAS*, 416(2):1443–1455, 2011.

[6] F. Pont, H. Knutson, R.L. Gilliland, C. Moutou, and D. Charbonneau. Detection of atmospheric haze on an extrasolar planet: the 0.55–1.05 μm transmission spectrum of HD 189733b with the Hubble Space Telescope. *MNRAS*, 385(1):109–118, 2008.

[7] D.K. Sing, C.M. Huitson, M. Lopez-Morales, F. Pont, J.M. Désert, D. Ehrenreich, P.A. Wilson, G.E. Ballester, J.J. Fortney, A.L. des Etangs, et al. Gtc osiris transiting exoplanet atmospheric survey: detection of sodium in XO-2b from differential long-slit spectroscopy. *MNRAS*, 426(2):1663–1670, 2012.

[8] N.P. Gibson, S. Aigrain, J.K. Barstow, T.M. Evans, L.N. Fletcher, and P.G.J. Irwin. A gemini ground-based transmission spectrum of WASP-29b: a featureless spectrum from 515 to 720 nm. *MNRAS*, page sts307, 2012.

[9] N.P. Gibson, S. Aigrain, J.K. Barstow, T.M. Evans, L.N. Fletcher, and P.G.J. Irwin. The optical transmission spectrum of the hot Jupiter HAT-P-32b: clouds explain the absence of broad spectral features? *MNRAS*, 436(4):2974–2988, 2013.

[10] A. Burrows, I. Hubeny, J. Budaj, H.A. Knutson, and D. Charbonneau. Theoretical spectral models of the planet HD 209458b with a thermal inversion and water emission bands. *ApJ*, 668(2):L171, 2007.

[11] H.A. Knutson, D. Charbonneau, A. Burrows, F.T. O'Donovan, and G. Mandushev. Detection of a temperature inversion in the broadband infrared emission spectrum of TrES-4. *ApJ*, 691(1):866, 2009.

[12] I. Hubeny, A. Burrows, and D. Sudarsky. A possible bifurcation in atmospheres of strongly irradiated stars and planets. *ApJ*, 594(2):1011, 2003.

[13] C. Marois, B. Macintosh, T. Barman, B. Zuckerman, I. Song, J. Patience, D. Lafrenière, and R. Doyon. Direct imaging of multiple planets orbiting the star HR 8799. *Science*, 322(5906):1348–1352, 2008.

[14] D. Stello, W.J. Chaplin, H. Bruntt, O.L. Creevey, A. García-Hernández, M. Monteiro, A. Moya, P.O. Quirion, S.G. Sousa, J.C. Suárez, et al. Radius determination of solar-type stars using asteroseismology: What to expect from the kepler mission. *ApJ*, 700(2):1589, 2009.

[15] H. Knutson, N. Madhusudhan, N. Cowan, J. Christiansen, E. Agol, D. Deming, J.M. Désert, D. Charbonneau, G. Henry, D. Homeier, et al. A Spitzer transmission spectrum for the exoplanet GJ436b, evidence for stellar variability, and constraints on dayside flux variations. *ApJ*, 735(1):27, 2011.

[16] J.M. Désert, A. Lecavelier Des Etangs, G. Hébrard, D. Sing, D. Ehrenreich, R. Ferlet, and A. Vidal-Madjar. Search for carbon monoxide in the atmosphere of the transiting exoplanet HD 189733b. *ApJ*, 699(1):478, 2009.

[17] J.J. Fortney, K. Lodders, M.S. Marley, and R.S. Freedman. A unified theory for the atmospheres of the hot and very hot jupiters: Two classes of irradiated atmospheres. *ApJ*, 678:1419, 2008.

[18] G.R. Ricker, J.N. Winn, R. Vanderspek, D.W. Latham, G.. Bakos, J.L. Bean, Z.K. Berta-Thompson, T.M. Brown, L. Buchhave, N.R. Butler, et al. Transiting exoplanet survey satellite (tess). *Proc. SPIE*, 9143:914320–914320–15, 2014.

[19] C. Hellier, D. Anderson, A. Collier Cameron, M. Gillon, L. Hebb, P. Maxted, D. Queloz, B. Smalley, A. Triaud, R. West, et al. An orbital period of 0.94 days for the hot-jupiter planet WASP-18b. *Nature*, 460(7259):1098–1100, 2009.

[20] D.J.A. Brown, A. Collier Cameron, C. Hall, L. Hebb, and B. Smalley. Are falling planets spinning up their host stars? *MNRAS*, 415(1):605–618, 2011.

[21] S. Nymeyer, J. Harrington, R. Hardy, K. Stevenson, C. Campo, N. Madhusudhan, A. Collier-Cameron, T. Loredo, J. Blecic, W. Bowman, C. Britt, P. Cubillos, C. Hellier, M. Gillon, P. Maxted, L. Hebb, P. Wheatley, D. Pollacco, and D. Anderson. Spitzer secondary eclipses of WASP-18b. *ApJ*, 742:35, 2011.

[22] P.F.L. Maxted, C. Koen, and B. Smalley. (ubv (ri) _c photometry of transiting planet host stars). *MNRAS*, 418:1039, 2011.

[23] A. Dressler, B. Bigelow, T. Hare, B. Sutin, I. Thompson, G. Burley, H. Epps, A. Oemler Jr, A. Bagish, C. Birk, et al. IMACS: The Inamori-Magellan areal camera and spectrograph on Magellan-Baade. *Publications of the Astronomical Society of the Pacific*, 123(901):288–332, 2011.

[24] R.M. Cutri, M.F. Skrutskie, S. Van Dyk, C.A. Beichman, J.M. Carpenter, T. Chester, L. Cambresy, T. Evans, J. Fowler, J. Gizis, et al. 2mass all sky catalog of point sources. 2003.

[25] F. Pont, S. Zucker, and D. Queloz. The effect of red noise on planetary transit detection. *MNRAS*, 373(1):231–242, 2006.

[26] O. Tamuz, T. Mazeh, and S. Zucker. Correcting systematic effects in a large set of photometric light curves. *MNRAS*, 136:1466, 2005.

[27] D.L. Pollacco, I. Skillen, A. Collier Cameron, D.J. Christian, C. Hellier, J. Irwin, T.A. Lister, R.A. Street, Richard G. West, D. Anderson, et al. The WASP project and the SuperWASP cameras. *Publications of the Astronomical Society of the Pacific*, 118(848):1407–1418, 2006.

[28] J. Bean, E. Kempton, and D. Homeier. A ground-based transmission spectrum of the super-Earth exoplanet GJ 1214b. *Nature*, 468(7324):669–672, 2010.

[29] P.A. Wilson, D.K. Sing, N. Nikolov, A. Lecavelier des Etangs, F. Pont, J.J. Fortney, G.E. Ballester, M. López-Morales, J.M. Désert, and A. Vidal-Madjar. GTC OSIRIS transiting exoplanet atmospheric survey: detection of potassium in HAT-P-1b from narrow-band spectrophotometry. *MNRAS*, 450(1):192–200, 2015.

[30] J. Eastman, B.S. Gaudi, and E. Agol. EXOFAST: A fast exoplanetary fitting suite in IDL. *Publications of the Astronomical Society of the Pacific*, 125(923):83–112, 2013.

[31] A. Triaud, A. Collier Cameron, D. Queloz, D.R. Anderson, M. Gillon, L. Hebb, C. Hellier, B. Loeillet, P.F.L. Maxted, M. Mayor, et al. Spin-orbit angle measurements for six southern transiting planets-new insights into the dynamical origins of hot jupiters. *Astronomy & Astrophysics*, 524:A25, 2010.

[32] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

[33] K. Mandel and E. Agol. Analytic light curves for planetary transit searches. *ApJ*, 580(2):L171, 2002.

[34] J. Southworth. Homogeneous studies of transiting extrasolar planets - I. light-curve analyses. *MNRAS*, 386(3):1644–1666, 2008.

[35] A. Claret and S. Bloemen. Gravity and limb-darkening coefficients for the Kepler, CoRoT, Spitzer, uvby, UBVRIJHK, and Sloan photometric systems. *Astronomy & Astrophysics*, 529:A75, 2011.

[36] D.R. Anderson, A. Collier Cameron, C. Hellier, M. Lendl, T.A. Lister, P.F.L. Maxted, D. Queloz, B. Smalley, A.M.S. Smith, A. Triaud, et al. WASP-31b: a low-density planet transiting a metal-poor, late-f-type dwarf star. *Astronomy & Astrophysics*, 531:A60, 2011.

[37] R.G. West, J.M. Almenara, D.R. Anderson, F. Bouchy, D.J.A. Brown, A. Collier Cameron, M. Deleuil, Laetitia. Delrez, A.P. Doyle, F. Faedi, et al. Three irradiated and bloated hot jupiters: WASP-76b, WASP-82b & WASP-90b. *arXiv preprint arXiv:1310.5607*, 2013.

A Sys-Rem Code

```
from scipy.optimize import curve_fit
import numpy as np
from math import log10, sqrt
from pylab import *
import matplotlib.pyplot as plt

##          0        1        2        3        4        5
stars = ['ob001', 'ob004', 'ob029', 'ob009', 'ob018', 'ob035']
labels = ['1' ,'4', '29', '9', '18', '35', 't']

## import data saved in text files ---
x = np.genfromtxt("data_1.txt", dtype=None)
y = np.genfromtxt("data_2.txt", dtype=None)
z = np.genfromtxt("data_3.txt", dtype=None)

flux1 = [0]*6
flux2 = [0]*6
flux3 = [0]*6

for i in range (0,6):    # star fluxes
    flux1[i] = x[:,i]
    flux2[i] = y[:,i]
    flux3[i] = z[:,i]

t1 = x[:,6]      # time of exposure
t2 = y[:,6]
t3 = z[:,6]

et1 = x[:,7]     # exposure times
et2 = y[:,7]
et3 = z[:,7]

a_j1 = x[:,8]      # airmass
a_j2 = y[:,8]
a_j3 = z[:,8]

m_set = [508, 313, 607]
flux = [flux1, flux2, flux3]
et = [et1, et2, et3]
a_j = [a_j1, a_j2, a_j3]
```

```
t = [t1, t2, t3]

## choose dataset and create terms for Sys-Rem ---
ds = 1  # choose which dataset to use - 1, 2 or 3
N = 6   # number of stars - range of i
M = m_set[ds - 1]     # number of measurements - range of j
flux1 = flux[ds-1]
et1 = et[ds-1]
a_j1 = a_j[ds-1]     # start with this air mass
t1 = t[ds-1]

start = [65, 18, 148]   # start of transit
end = [307, 181, 387]   # end of transit
start = start[ds-1]
end = end[ds-1]

flfl = []
for j in range(M):
    flfl.append(flux1[0][j])

mag1 = [0]*N
ave1 = [0]*N
mag_ave1 = [0]*N
r_ij = [0]*N
sig_ij = [0]*N
for i in range(0,N):
    mag1[i] = [0]*M
    r_ij[i] = [0]*M
    sig_ij[i] = [0]*M

for i in range(N):
    for j in range(M):
        flux1[i][j] = flux1[i][j]/et1[j]     # convert counts to
            fluxes
mag1=flux1  # use flux

for i in range(N):
    mag_ave1[i] = np.average(mag1[i])
    for j in range(M):
        r_ij[i][j] = mag1[i][j] - mag_ave1[i]    # residual of each
            observation
        sig_ij[i][j] = sqrt(flux1[i][j])     # use flux
```

```
## choose stars to use (0 is WASP-18)
ss = [0,1,2,3,4,5] #[0,1,2,3,4,5]

# change arrays to only have stars wanted
xyz = [0]*N
for i in ss:
    xyz[i] = []
    xyz[i].append(mag1[i])      # [i][0]
    xyz[i].append(r_ij[i])
    xyz[i].append(sig_ij[i])
    xyz[i].append(mag_ave1[i])

mag1 = []
r_ij = []
sig_ij = []
mag_ave1 = []
for i in ss:
    if xyz[i] != 0:
        mag1.append(xyz[i][0])
        r_ij.append(xyz[i][1])
        sig_ij.append(xyz[i][2])
        mag_ave1.append(xyz[i][3])
    else:
        print 'Not using star number', i

N = len(ss)      # number of stars chosen

if ds == 2:      # cut bad points from start of night 2
    M = M-6
    mag1 = mag1[6:]
    flfl = flfl[6:]
    et1 = et1[6:]
    a_j1 = a_j1[6:]
    t1 = t1[6:]
    for i in range(N):
        r_ij[i] = r_ij[i][6:]
        sig_ij[i] = sig_ij[i][6:]
        flux1[i] = flux1[i][6:]

## use comparisons many times each ---
```

```
ns = 50  # number of times to use comparison stars
r_ij_e = []
sig_ij_e = []
flux1_e = []
mag_ave1_e = []

for k in range(ns-1):
    for i in range(len(ss)):
        if i != 0:
            r_ij_e.append(r_ij[i])
            sig_ij_e.append(sig_ij[i])
            flux1_e.append(flux1[i])
            mag_ave1_e.append(mag_ave1[i])

r_ij = r_ij + r_ij_e
sig_ij = sig_ij + sig_ij_e
flux1 = flux1 + flux1_e
mag_ave1 = mag_ave1 + mag_ave1_e

N = len(r_ij)

## Sys-Rem code ---
loops = 0
numtrends = 1   # number of times to run sysrem (i.e. number of
   trends to look for)
while loops < numtrends:

    ## find a and c that minimise S
    c_loop = []      # holds c's found in each loop
    a_loop = []      # holds a's found in each loop
    numloops = 0
    numend = 50  # number of times to iterate to find a & c
    c_i_ar = [0]*N

    while numloops < numend:

        for i in range(N):     # find c from a
            eq1 = []
            eq2 = []
            for j in range(M):  # do calculation inside sum
                eq1.append(r_ij[i][j] * a_j1[j] / ((sig_ij[i][j])
                   **2.0))
```

```
            eq2.append(a_j1[j] * a_j1[j] / ((sig_ij[i][j])
                **2.0))
        ctop = sum(eq1)
        cbottom = sum(eq2)
        c_i_ar[i] = ctop / cbottom
    c_loop.append(c_i_ar)

    for j in range(M):  # find a from c
        eq3 = []
        eq4 = []
        for i in range(N):
            eq3.append(r_ij[i][j] * c_i_ar[i] / (sig_ij[i][j])
                **2.0)
            eq4.append(c_i_ar[i] * c_i_ar[i] / (sig_ij[i][j])
                **2.0)
        atop = sum(eq3)
        abottom = sum(eq4)

        a_j1[j] = atop / abottom      # for next loop
    a_loop.append(a_j1)

    numloops += 1

#print a_j1  # final values
#print c_i_min

c_i = c_i_ar

## remove trend found
r_ij_new = [0]*N    # calculate new residuals
for i in range(N):
    r_ij_new[i] = [0]*M
    for j in range(M):
        r_ij_new[i][j] = r_ij[i][j] - c_i[i] * a_j1[j]

new = [0]*N      # corrected flux
for i in range(N):
    new[i] = [0]*M
    for j in range(M):
        new[i][j] = mag_ave1[i] + r_ij_new[i][j]

r_ij = r_ij_new      # for next trend (next run of sysrem)
```

```
    loops = loops + 1
    #print loops, '/', numtrends

    testa = set(a_loop[-2]) == set(a_loop[-1])  # tests to see if
       converged on a and c
    testc = set(c_loop[-2]) == set(c_loop[-1])
    if testa is False or testc is False:
        print '>>> a or c arrays different still'

f_b = flux1      # flux before sysrem
f_a = new      # flux after sysrem

##plot(f_b[0], marker='.', linestyle='none', color='blue')  # plot
    before
##plot(f_a[0], marker='.', linestyle='none', color='red')     #
   plot after
##title('Before (blue) and after (red) Sys-Rem for transit')
##show()

##plot(f_b[1], marker='.', linestyle='none', color='blue')  # plot
    before
##plot(f_a[1], marker='.', linestyle='none', color='red')     #
   plot after
##title('Before (blue) and after (red) Sys-Rem for a single star')
##show()

for i in range(len(ss)):
    x = f_a[i]
    x = x / np.average(x) + i/50.0
    plot(t1, x,label=i, marker='.', linestyle='none')
title('All stars after Sys-Rem', fontsize=25)
xlabel('Julian date (days)', fontsize=15)
ylabel('Normalised flux', fontsize=15)
legend()
show()

## END
```

B Polynomial Detrending Code

```
import cPickle as pickle
import numpy as np
```

```
from pylab import *
from scipy.optimize import curve_fit
from math import log10, sqrt
import matplotlib.pyplot as plt
from astropy.stats import sigma_clip

LC1 = pickle.load(open('LCs_w18_all.pkl', 'r'))      # load data

ds = 1  # dataset number
Mset = [508,313,607]     # number of images
M = Mset[ds-1]
start = [65, 18, 148]    # start of transit
end = [307, 181, 387]    # end of transit
start = start[ds-1]
end = end[ds-1]

t = LC1['t'] + 2.4e6 +0.5    # convert time to BJD
wbins = LC1['wbins']         # wavelength bins
et = LC1['etimes']           # exposure times

olcw = LC1['oLCw']  # target light curves
clcw = LC1['cLCw']  # comparison light curves

## Binning ---
ob = zip(*olcw)  # bin-images [63][508]
cb = [0]*len(wbins)       # comp 1 only
for i in range(len(wbins)):
    cb[i] = [0]*M
    for j in range(M):
        cb[i][j] = clcw[j][0][i]

ob = np.asarray(ob)
cb = np.asarray(cb)

## Detrending for all bins ---
for bn in range(len(wbins)):
    try:
        F = ob[bn]/cb[bn]    # polynomial divide step

        Fout = []    # out-of-transit arrays
        tout = []
        for j in range(M):
```

```
    if j <= start or j >=end:
        Fout.append(F[j])
        tout.append(t[j])

Fout = np.asarray(Fout)
tout = np.asarray(tout)

td = []      # time from start in days
for j in range(M):
    td.append(t[j] - t[0])
for j in range(M):  # convert to hours
    td[j] = td[j] * 24.0

mid = int((start + end)/2.0)     # mid-point of transit

thc = []     # hours from mid-transit
for j in range(M):
    thc.append(td[j] - td[mid])

thcout = [] # time out-of-transit only
for j in range(M):
    if j <= start or j >=end:
        thcout.append(thc[j])
thcout = np.asarray(thcout)

func2 = []  # 2nd order fit
coeff, covar = curve_fit(lambda x, a, b, c: a*x**2.0 + b*x
    + c, thcout, Fout)
for j in range(M):
    func2.append(coeff[0]*thc[j]**2.0 + coeff[1]*thc[j] +
        coeff[2])

out = list(F[:start]) + list(F[end:])    # out-of-transit
    flux
print bn, np.average(out)
F_cor2 = F - func2 + np.average(out)     # correct data
    with polynomial

# error analysis
sig = [[],[]]
for j in range(M):
    sig[0].append(1.0/sqrt(ob[bn][j]))
```

```
            sig[1].append(1.0/sqrt(cb[bn][j]))

        sig_F = []
        for j in range(M):
            sig_F.append(sqrt( (sig[0][j]/cb[bn][j])**2.0 + (sig
                [1][j]**2.0) * (ob[bn][j]/cb[bn][j]**2.0)**2.0 ))
        for j in range(M):
            sig_F[j] = sig_F[j] * ob[bn][j]/cb[bn][j]    # convert
                relative to absolute

        # write to file for EXOFAST
        with open('final_bin_all_'+str(bn)+'.flux','w') as f1:
            lis1 = [t, F_cor2, sig_F]
            for x in zip(*lis1):
                f1.write("{0}\t{1}\t{2}\n".format(*x))

        # plot each detrended bin
        plot(t,F_cor2, linestyle='none', marker='.')
        xlabel('Time (BJD)')
        ylabel('Flux')
        title('Corrected Ft/Fc for bin '+str(bn))
        savefig('final_bin_'+str(bn))
        clf()

    except (RuntimeError, ValueError, ZeroDivisionError):
        print '>  Bin', bn, 'failed'

## END
```

C FWHM Code

```
import pyfits
import glob
from pylab import *
from scipy.optimize import curve_fit
import numpy as np
import matplotlib.pyplot as plt

# python function to return index of a value in an array
def fn(array,value):
    idx = (np.abs(array-value)).argmin()
    return idx
```

```
# load residuals of WLC
res = np.loadtxt('res_WLC.txt')

chip5 = glob.glob('ift*c5.fits')     # import names of files for
   chip 5
start = 800      # start y pixel of cut
stop = 830       # end y pixel of cut

fwhm = []
iar = []     # image number on file
rescut = []      # cut errors out
i_notused = []

for i in range(116,624):     # loop over images
    d,h = pyfits.getdata(chip5[i],header=True)     # data [y,x],
       header

    cut = d[1653][start:stop]

    pix = []
    for j in range(start,stop):
        pix.append(j-start)      # convert 800 -> 0 for curve fit

    func = []
    coeff, covar = curve_fit(lambda x, a, b, c, d: a*np.exp(-(x-c)
       **2.0/(2.0*b**2.0)) + d, pix, cut) # gaussian
    for j in range(len(pix)):
        func.append(coeff[0]*np.exp(-(pix[j]-coeff[2])**2.0/(2.0*
           coeff[1]**2.0)) + coeff[3])

    pixext = []
    dens = 150  # get better resolution of gaussian
    for j in range(dens*start, dens*stop):
        pixext.append((j-start*dens)/float(dens))
    funcext = []
    for j in range(len(pixext)):
        funcext.append(coeff[0]*np.exp(-(pixext[j]-coeff[2])
           **2.0/(2.0*coeff[1]**2.0)) + coeff[3])

##     if i == 300 or i == 400:     # plot of 2 cuts
##         plot(pixext, funcext)
##         plot(pix, cut, linestyle='none', marker='o')
```

```
##          minorticks_on()
##          plt.tick_params(which='both', width=1)
##          plt.tick_params(which='major', length=14)
##          plt.tick_params(which='minor', length=8)
##          xlabel('Spatial pixel number', fontsize=15)
##          ylabel('Counts', fontsize=15)
##          show()

    # get fwhm by hand ---
    hmax = max(funcext)/2.0      # half-maximum of fitted function
    try:
        mx = np.where(funcext == max(funcext))  # index of maximum

        # indexes of half-maximum either side of maximum
        indl = fn(funcext[:mx[0][0]], hmax)
        indh = fn(funcext[mx[0][0]:], hmax) + len(funcext[:mx
            [0][0]])

        iar.append(i)
        fwhm.append(pixext[indh] - pixext[indl])     # compute FWHM
        rescut.append(res[i-117])

    except ValueError:
        print '>  - Error for', chip5[i]
        i_notused.append(i)

    if i == 99+117 or i == 199+117 or i == 299+117 or i == 399+117
        or i == 499+117 or i == 599+117 or i == 699+117:
        print '>  Done image number =', i-116, '/', 508      #
            progress

plot(iar, fwhm, linestyle='none', marker='.')
title('FWHM for each image')
xlabel('Image number')
ylabel('FWHM (pixels)')
minorticks_on()
plt.tick_params(which='both', width=1)
plt.tick_params(which='major', length=14)
plt.tick_params(which='minor', length=8)
show()

fwhm = np.asarray(fwhm)
```

```
rescut = np.asarray(rescut)

func2 = []  # fit line to trend
##coeff2, covar2 = curve_fit(lambda x, a, b, c, d: a*x - b*x + c -
    d, fwhm, rescut)     # straight line
coeff2, covar2 = curve_fit(lambda x, a, b: a*x + b, fwhm, rescut)
       # straight line
for j in range(len(fwhm)):
##     func2.append(coeff2[0]*fwhm[j] - coeff2[1]*fwhm[j] + coeff2
    [2] - coeff2[3])
     func2.append(coeff2[0]*fwhm[j] + coeff2[1])

print coeff2

plot(fwhm, rescut, linestyle='none', marker='.')
plot(fwhm, func2)
title('FWHM against residuals')
xlabel('FWHM (pixels)')
ylabel('Residual')  # data - fit
minorticks_on()
plt.tick_params(which='both', width=1)
plt.tick_params(which='major', length=14)
plt.tick_params(which='minor', length=8)
show()

plot(fwhm[307:], rescut[307:], linestyle='none', marker='.')
##plot(fwhm[307:], func2[307:])
plot(fwhm[:65], rescut[:65], linestyle='none', marker='.')
##plot(fwhm[:65], func2[:65])
title('FWHM against residuals - out of transit only')
xlabel('FWHM (pixels)')
ylabel('Residual')  # data - fit
minorticks_on()
plt.tick_params(which='both', width=1)
plt.tick_params(which='major', length=14)
plt.tick_params(which='minor', length=8)
show()

## try for bins ---
i_notused.reverse()
print i_notused
```

```
bu = np.loadtxt('bins_used.txt')
resb = np.loadtxt('residuals.txt')
for i in range(len(bu)):
    res = list(resb[i])

    for j in i_notused:
        del(res[j])

    plot(fwhm, res, linestyle='none', marker='.')
    title('Bin '+str(bu[i]))
    xlabel('FWHM (pixels)')
    ylabel('Residual')  # data - fit
    savefig('res_bin_'+str(bu[i])+'.png')
    clf()

## END
```